

Teoria da Credibilidade: Uma Abordagem Bayesiana para Estimação de Prêmios de Risco em Seguros de Vida.

Autor: Juscelino Zemiacki

Orientadora: Prof. Dra. Patrícia K. Ziegelmann

Co-Orientador: Prof. Sérgio Rangel Guimarães

Porto Alegre, 13 de Dezembro de 2006.

Universidade Federal do Rio Grande do Sul
Instituto de Matemática
Departamento de Estatística

Teoria da Credibilidade: Uma Abordagem Bayesiana para Estimação de Prêmios de Risco em Seguros de Vida.

Autor: Juscelino Zemiacki

Monografia apresentada para obtenção
do grau de Bacharel em Estatística.

Banca Examinadora:
Professora Dra. Patrícia Klarmann Ziegelmann (orientador)
Professor Dr. Lorí Viali (convidado do Departamento de Estatística da UFRGS)

Porto Alegre, 13 de Dezembro de 2006.

AGRADECIMENTOS

Uma vez encerrada uma etapa em nossa vida, o momento mais importante é, sem dúvida, o momento de agradecer. Agradecer, porque sozinho dificilmente alguém consegue realizar e percorrer um caminho até o fim.

Agradeço a Deus, fonte de fé e força em todos os momentos.

Agradeço a minha família, razão maior da minha vida: meu pai André, minha mãe Iraci, meu irmão Leonel Paulo, minha irmã Maria Izabel. Meus pais, obrigado por serem meus guias e por terem feito com que eu entendesse o verdadeiro sentido do caminho correto para a vida.

Agradeço ao grande amor da minha vida, Eunice, por fazer com que eu me sinta uma pessoa tão especial e capaz de vencer a qualquer obstáculo da minha vida. "*Meine Liebe, ich liebe dich*"!

Agradeço a professora Patrícia K. Ziegelmann, minha orientadora. Obrigado pela sua dedicação, paciência, orientação e incentivo, não só na realização deste trabalho, mas também ao longo do curso, nas disciplinas em que tive o privilégio de tê-la como professora. Muito obrigado por tudo!

Ao professor Sérgio Rangel, que me orientou na escolha do tema deste trabalho. Obrigado pela orientação e colaboração.

Agradeço a todos meus mestres do curso de Estatística pela orientação e pelos ensinamentos, muito úteis na minha formação, não só acadêmica como também humana.

Agradeço a professora Ana Maria Escobar Bernardes, pela oportunidade de meu primeiro estágio, tão importante para o meu crescimento e minha formação. Muito obrigado pelos seus ensinamentos e por ter assumido papel de mãe em muitas oportunidades.

Ao colega de casa do estudante Jonas, futuro Estatístico, que me auxiliou em trabalhos profissionais e permitiu que eu tivesse tempo para concluir essa monografia.

E por fim, agradeço a todas as pessoas que de uma forma ou outra colaboraram e estiveram presentes ao longo dessa longa caminhada. Em especial as colegas e amigas: Sabrina, Adriana e Marilyn. Obrigado pelas horas de estudo compartilhadas. Em especial também aos quase irmãos da Casa do estudante Universitário: Liana, Gregório, André, César, Raquel, Mauricio, Dione. Obrigado pela recepção, amizade e por terem feito com que me sentisse em casa.

RESUMO

A atividade seguradora, ou simplesmente seguro, surgiu como um mecanismo criado a partir da necessidade do ser humano em buscar proteção, para si ou para seu patrimônio, em relação ao acaso, o qual, quando implica em prejuízos econômicos, é denominado de Risco, na Ciência Atuarial. Assim, o seguro constitui-se em uma forma de restabelecimento de um equilíbrio que foi perturbado pela materialização ou concretização do risco. Buscando-se estabelecer a equivalência entre o prêmio cobrado do segurado e o risco assumido pelo segurador surgiram diversas técnicas na área de tarifação de seguros, cujos conceitos, metodologias e princípios envolvidos nos cálculos para tarifação do prêmio de seguro, se baseiam na lógica elementar de que o valor cobrado deve cobrir os custos inerentes ao risco assumido. Neste contexto, a Teoria da Credibilidade é apontada pela literatura atuarial como uma das técnicas mais inovadoras e avançadas no processo de previsão, tarefa essa imprescindível para a tarifação de prêmios de seguros. Além de sua aplicação nessa área, esta técnica possui uma grande diversidade de aplicações, tais como, resseguros, definição de reservas de sinistros e até mesmo na calibração de instrumentos: como tábuas biométricas e curvas de mortalidade. O objetivo deste trabalho é demonstrar a aplicação dos métodos Bayesianos na Teoria da Credibilidade, através de uma aplicação prática em seguros de vida a partir de dados reais da Seguradora Previdência do Sul. Visando este objetivo, apresentam-se neste trabalho as principais definições e conceitos Bayesianos, assim como suas aplicações aos modelos de Credibilidade de Bühlmann e Bühlmann-Straub, através da implementação dos métodos Bayesianos com uso da técnica de simulação Gibbs Sampling. A aplicação da Teoria da Credibilidade sob o enfoque Bayesiano em dados reais de uma apólice de seguro de vida se mostrou realmente efetiva em seu intuito de prever o comportamento futuro de uma determinada variável utilizada na obtenção do prêmio de risco, podendo assim ser aplicada em diversas situações práticas.

ÍNDICE

1. Introdução	01
2. Fundamentos Bayesianos para a teoria da Credibilidade	06
2.1. Teorema de Bayes	08
2.2. Distribuição a Priori e Distribuição a Posteriori	09
2.3. Priori Conjugada	11
2.3.1. Priori Não-informativa.....	12
2.4. Resultados a Posteriori	13
2.5. Distribuição Preditiva a Posteriori	15
2.6. Métodos de Monte Carlo baseados em Cadeias de Markov	15
2.6.1. Algoritmo de Metropolis-Hastings	17
2.6.2. Algoritmo Gibbs Sampling	19
2.6.3. A amostra da Posteriori	20
3. Teoria da Credibilidade	22
3.1. Teoria da Credibilidade de Maior Exatidão	23
3.2. Tamanho da Experiência	27
3.3. Modelo de Bühlmann	29
3.3.1. Abordagem Clássica para o Modelo de Bühlmann	31
3.3.2. Abordagem Bayesiana para o Modelo de Bühlmann	32
3.4. Modelo de Bühlmann-Straub	34
3.4.1. Abordagem Clássica do Modelo de Bühlmann-Straub	34
3.4.2. Abordagem Bayesiana para o Modelo de Bühlmann-Straub ..	36
3.5. Simulação Aplicada ao Modelo de Bühlmann	36

4. Exemplo prático aplicado em Seguros de vida em grupo.....	42
4.1. Descrição do banco de dados	44
4.2. Procedimentos utilizados para análise	45
4.3. Análises e resultados	46
5. Considerações finais	56
Referências Bibliográficas	58
Anexo 1 – Definições básicas de Cadeias de Markov	59
Anexo 2 – Programação em R para exemplo de simulação	62
Anexo 3 – Programação no Winbuggs para exemplo prático	66

1. INTRODUÇÃO

A atividade seguradora, ou simplesmente seguro, surgiu como um mecanismo criado a partir da necessidade do ser humano em buscar proteção, para si ou para seu patrimônio, em relação ao acaso. Este acaso, quando implica em prejuízos econômicos é classificado, na Ciência Atuarial, como Risco. Assim, o seguro se constitui como uma forma de restabelecimento de um equilíbrio que foi perturbado pela materialização ou concretização do risco.

São elementos essenciais do seguro - o segurador (ou empresa seguradora), o segurado, o prêmio e o risco, os quais são definidos a seguir.

- Seguradora: instituição que tem por objetivo indenizar prejuízos involuntários verificados no patrimônio de outrem, mediante recebimento de prêmios.

- Segurado: pessoa física ou jurídica que, tendo interesse segurável, contrata o seguro, em seu benefício pessoal ou de terceiros.

- Prêmio: a importância paga pelo segurado, ou estipulante, ao segurador em troca da transferência do risco a que ele está exposto.

- Risco: evento possível, com ocorrência e momento de ocorrência incerta, independente da vontade das partes contratantes e contra o qual é feito o seguro. O risco é a expectativa de sinistro. Sem risco não pode haver contrato de seguro. É comum a palavra ser usada, também, para significar a coisa ou pessoa sujeita ao risco.

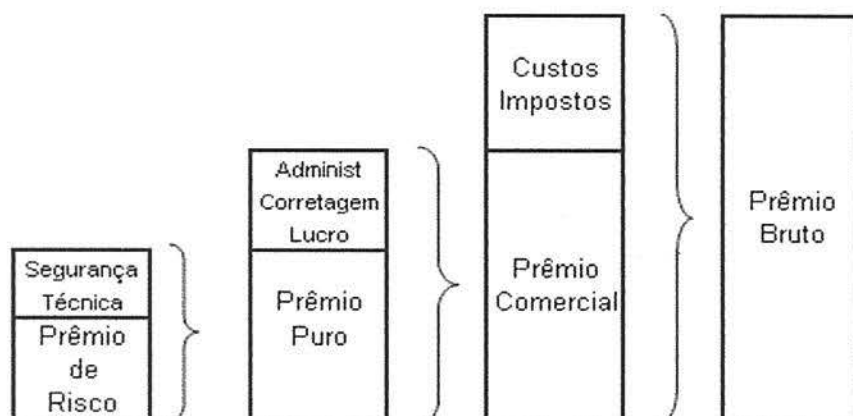
De maneira geral, pode-se dizer que a atividade seguradora consiste na administração de dois fluxos de capital: o prêmio pago pelo segurado ao segurador e a indenização paga pelo segurador ao segurado. O esquema abaixo mostra a relação entre estes dois fluxos.



Buscando-se estabelecer o equilíbrio entre o prêmio cobrado e o risco assumido pelo segurador surgiram diversas técnicas na área de tarifação de seguros, cujos conceitos e metodologias envolvidos nos cálculos para precificação do prêmio de seguro, assim como também os princípios dos cálculos de prêmios, se baseiam na lógica elementar de que o valor cobrado deve cobrir os custos inerentes ao risco assumido. Assim sendo, a tarifação pode ter seu significado entendido como métodos de obtenção de prêmios a serem pagos pelo segurado à seguradora, de modo que esta possa cumprir o compromisso de pagar indenizações ao segurado, de acordo com a frequência esperada de sinistros.

Em um processo de tarifação do custo de um seguro, uma das formas de classificação dos prêmios é dada em função da composição de seus custos, os quais podem

ser desdobrados em: prêmio de risco, prêmio puro, prêmio comercial e prêmio bruto, conforme ilustração a seguir.



O prêmio de risco indica na sua essência, a esperança matemática de sinistros futuros; o prêmio puro se obtém ao agregar no prêmio de risco uma margem, denominada carregamento técnico de segurança; já o prêmio comercial é o prêmio puro acrescido dos custos administrativos, custos de corretagem e margem de lucro; Ainda há o prêmio bruto, o qual é resultante do prêmio comercial acrescido dos impostos que incidem sobre o custo da apólice. Desta forma, a obtenção do prêmio de risco pode ser considerada como uma etapa fundamental na tarificação de um prêmio de seguro, a exceção de situações onde o prêmio é pré-determinado por entidades governamentais supervisoras da atividade seguradora. Além disso, pode-se dizer que o prêmio de risco está diretamente ligado a uma variável aleatória que espelhe o prêmio a ser taxado, ou seja, está diretamente sofrendo efeitos do acaso.

Os métodos de tarificação para a obtenção do prêmio de risco são praticados basicamente sob dois enfoques distintos, sendo eles: tarificação por experiência e tarificação por atributos. A tarificação por experiência se baseia em dados observados das carteiras ou

apólices estudadas, ou seja, na experiência observada do grupo de segurados dessas carteiras. Enquanto a tarifação por atributos, se baseia em conhecimentos subjetivos do atuário, fundamentado em sua própria experiência em relação às características do segurado, ou em conhecimentos padrão, como por exemplo, tábuas Biométricas.

Neste contexto, aplica-se a Teoria da Credibilidade, que, sob um enfoque Bayesiano, pode ser entendida como uma técnica de tarifação que pode unir estes dois tipos de tarifação: tarifação por experiência e tarifação por atributos, ou seja, o conhecimento subjetivo do atuário com os dados observados do grupo de segurados. Foi exatamente com esse sentido que a palavra credibilidade se originou na Ciência Atuarial. A palavra credibilidade se reportava a uma medida subjetiva atribuída pelo atuário a determinado fator de interesse com o objetivo de mensurar ou refletir sua experiência pessoal neste fator com vistas a tarifar, ou seja, definir os valores dos prêmios de seguro. Essa idéia casa perfeitamente com a Inferência Bayesiana, cuja principal característica é permitir que se agregue a análise informações alheias aos dados observados.

Atualmente a Teoria da Credibilidade é apontada pela literatura atuarial como uma das técnicas mais inovadoras e avançadas no processo de tarifação de seguros. Além de sua aplicação nessa área, esta técnica possui uma grande diversidade de aplicações, tais como, resseguros, definição de reservas de sinistros e até mesmo na calibração de instrumentos: como tábuas biométricas e curvas de mortalidade. Assim, devido à importância destacada a Teoria da Credibilidade na área de tarifação de seguros, a abordagem desta técnica neste trabalho objetiva, principalmente, ilustrar a utilização desta no ramo de seguros de vida já que grande parte dos trabalhos com aplicações da Teoria da Credibilidade existentes na literatura se restringe a ramos de seguro não-vida, como seguros de automóveis, por exemplo. Mais especificamente este trabalho tem como objetivo principal a apresentação e implementação de métodos Bayesianos para a Teoria da Credibilidade.

Com este objetivo, o trabalho será constituído e estruturado em mais quatro capítulos, de forma a apresentar os principais conceitos e fundamentações Bayesianas da

Teoria da Credibilidade, assim como uma ilustração prática da técnica, com utilização de dados reais de uma apólice de seguro de vida da Seguradora Previdência do Sul.

No Capítulo 2 apresentam-se os conceitos básicos de Inferência Bayesiana utilizados na Teoria da Credibilidade, bem como a descrição de técnicas computacionais de simulação de Monte Carlo através de Cadeias de Markov, com uma breve descrição dos algoritmos de Metropolis-Hasting e Gibbs Sampling.

No Capítulo 3 apresenta-se uma introdução a Teoria da Credibilidade, com um breve histórico dessa técnica. Faz-se uma demonstração rápida da derivação da fórmula da Teoria da Credibilidade e se apresenta os modelos de Bühlmann e Bühlmann-Straub, vistos sob o enfoque Clássico e sob o enfoque Bayesiano. Neste capítulo também se apresenta um exercício de simulação, abordando a discussão sobre o tamanho da experiência desejada para que uma apólice tenha credibilidade completa dentro de uma carteira de seguros.

No Capítulo 4 desenvolve-se um exemplo de aplicação prática da Teoria da Credibilidade Bayesiana em dados de uma apólice. Para isso se utilizou de dados de seguro de vida em grupo, da seguradora Previdência do Sul.

Por fim, no Capítulo 5 se apresentam as considerações finais e conclusões sobre este trabalho.

Além desses cinco capítulos, o trabalho possui ainda dois anexos. No anexo I se apresenta os conceitos básicos de cadeia de Markov utilizados em simulações Monte Carlo. No anexo II se apresenta os programas realizados em software R e Winbugs utilizados neste trabalho.

2. FUNDAMENTOS BAYESIANOS PARA A TEORIA DA CREDIBILIDADE

Atualmente, a Teoria da Credibilidade se embasa nos métodos da Inferência Bayesiana, pois estes permitem que informações subjetivas sobre o Risco analisado sejam incorporadas no modelo e, ainda, porque certas características inerentes a área de Tarifação de seguros também podem ser mais facilmente incorporadas aos modelos. A Teoria Bayesiana data de 1763, quando o reverendo Thomas Bayes, em seu artigo original, Bayes (1763), explicitou a relação conhecida como Teorema de Bayes. Os métodos Bayesianos consistem na fusão de uma informação inicial sobre o parâmetro desconhecido e que se quer estimar, expressada mediante uma distribuição de probabilidade conhecida como *priori*, com as observações estatísticas, os dados, para produzir uma informação final, representada por uma distribuição de probabilidade conhecida como *posteriori*, a qual sintetiza ambas as fontes de informações e é a base para obter conclusões e tomar decisões. Estas são as idéias básicas do Teorema de Bayes.

Na Inferência Clássica os parâmetros desconhecidos do modelo em estudo são considerados constantes. Já na Inferência Bayesiana estes parâmetros são consideradas quantidades aleatórias. Sendo assim é natural que, na abordagem Estatística Bayesiana, sejam atribuídas distribuições de probabilidade aos parâmetros. Esta é a principal diferença entre a Inferência Estatística Clássica e a Inferência Estatística Bayesiana, ou seja, a abordagem Bayesiana baseia-se em distribuições de probabilidade para os parâmetros desconhecidos. Mais especificamente, toda a inferência a cerca do parâmetro desconhecido do modelo a ser realizada sob o enfoque Bayesiano é baseada na distribuição a Posteriori para este parâmetro, a qual é uma combinação entre a informação a priori a cerca do parâmetro e as informações dos dados amostrais.

Assim, o principal objetivo da Inferência Bayesiana é obter a distribuição a posteriori. Para a obtenção dessas posteriores geralmente se faz necessário escolher distribuições a *priori* que facilitem os cálculos necessários. Neste sentido surge a idéia de utilização das chamadas *prioris* conjugadas ao modelo em estudo. No entanto, mesmo com

a utilização de prioris conjugadas, problemas mais complicados que requerem modelos mais complexos geralmente conduzem à *posterioris* analiticamente intratáveis. Para resolver estes problemas técnicas computacionais, como por exemplo, os Métodos Monte Carlo, em especial *Markov Chain Monte Carlo* (MCMC), surgem como alternativa. Em resumo, estas técnicas fornecem uma metodologia para a obtenção de amostras oriundas da posteriori de interesse mesmo quando sua derivação analítica é impraticável.

Na teoria da credibilidade, além de uma estimativa para θ (parâmetro de interesse, o qual pode ser número médio de sinistros, valor médio de indenizações pagas, taxas de risco, entre outros), busca-se prever o comportamento de uma realização futura dessa variável de interesse. O objetivo de predição enquadra-se perfeitamente nas idéias dos métodos Bayesianos através da chamada distribuição de probabilidade preditiva a posteriori.

Assim, neste capítulo, se objetiva definir e apresentar os principais elementos da Inferência Bayesiana aplicados na Teoria da Credibilidade. Na Seção 2.1 se apresenta a definição do Teorema de Bayes. Na Seção 2.2 se apresenta a definição de distribuições a priori e a posteriori. Já na Seção 2.3 se introduz a idéia de Prioris conjugadas, sendo estas exemplificadas através do modelo de Poisson e do modelo Normal. A Seção 2.4 apresenta as idéias sobre o processo inferencial, explicitando como estimativas pontuais e por intervalo são obtidas sob um enfoque Bayesiano. Na Seção 2.5 define-se distribuição preditiva a posteriori e, por último, na Seção 2.6 se apresentam os métodos MCMC, com uma breve descrição dos algoritmos de Metropolis-Hastings e Gibbs Sampling.

2.1. Teorema de Bayes

Seja $\{A_i\}$, $i=1, 2, \dots, n$, uma partição de um espaço amostral (Ω) . Então $\bigcup_{i=1}^n A_i = \Omega$

onde A_i pode ser interpretado como as possíveis causas de um evento B qualquer de Ω . Supondo que se conhece o evento B , se deseja a probabilidade de que este evento tenha sido causado especificamente por A_i . Aplicando-se a definição de probabilidade condicionada, (Meyer, 1983), essa probabilidade pode ser expressa como:

$$P(A_i / B) = \frac{P(A_i)P(B / A_i)}{\sum_{j=1}^n P(B / A_j)P(A_j)}, i=1, 2, \dots, n \quad (2.1)$$

Segundo Meyer, (1983), este resultado é conhecido como Teorema de Bayes, também denominado como fórmula da probabilidade das causas ou dos antecedentes. Isso porque, desde que os A_i constituam uma partição do espaço amostral, um e somente um dos eventos A_i ocorrerá, ou seja, um dos eventos A_i deverá ocorrer e somente um poderá ocorrer. Portanto, o Teorema de Bayes expressa a probabilidade de um particular A_i (isto é, uma causa) dado que o evento B tenha ocorrido. Para a aplicação do Teorema de Bayes se faz necessário conhecer as probabilidades de ocorrência dos A_i , ou seja, $P(A_i)$. Muito frequentemente essas probabilidades são escolhidas de forma a expressar o conhecimento *a priori* sobre a ocorrência do evento.

2.2. Distribuições a Priori e a Posteriori

Na Estatística Bayesiana a priori representa a informação prévia acerca do parâmetro desconhecido θ do modelo. Esta informação pode consistir no conhecimento de

algum especialista no assunto ou ainda em informações passadas. Assim, a escolha da Distribuição a Priori é de responsabilidade do analista ou pesquisador.

A posteriori é uma distribuição de probabilidade para o parâmetro θ que indica o conhecimento acerca do parâmetro depois de observada uma amostra \underline{x} . Ela conjuga as informações a priori com as informações dos dados de uma amostra representando o atual conhecimento sobre o parâmetro desconhecido do modelo. Assim, uma possível aplicação do teorema de Bayes, definido na Seção 2.1, é a que se apresenta a seguir.

Seja X_1, X_2, \dots, X_n uma amostra de variáveis aleatórias independentes e identicamente distribuídas. Considere que a função de probabilidade de cada variável X_i será dependente de um parâmetro desconhecido θ . No contexto Bayesiano θ , por ser desconhecido, é visto como quantidade aleatória, sendo assim natural a notação $f(x_i / \theta)$ para a função de probabilidade de cada X_i , ou seja, a distribuição de X_i é condicional ao parâmetro θ . Dado que X_1, X_2, \dots, X_n são independentes e identicamente distribuídas, a função de probabilidade condicional conjunta das variáveis aleatórias X_1, X_2, \dots, X_n é expressa por:

$$f(\underline{x} / \theta) = \prod_{i=1}^n f(x_i / \theta). \quad (2.2)$$

Fazendo conexão com o Teorema de Bayes, a amostra observada \underline{x} é o evento e o parâmetro θ a causa. Assim sendo, a idéia então é, dado que foi observada uma amostra \underline{x} , se deseja determinar a probabilidade de que essa amostra tenha sido causada especificamente por um determinado valor de θ . Ou seja, $P(\theta = \theta_i / \underline{X} = \underline{x})$, no caso de θ discreto e $f(\theta / \underline{X} = \underline{x})$, no caso de θ contínuo. Para facilitar a notação será utilizado $\pi(\theta / \underline{x})$ para ambos os casos.

Pelo Teorema de Bayes se tem que:

$$\pi(\theta/\underline{x}) = \frac{f(\underline{x}/\theta)\pi(\theta)}{\int f(\underline{x}/\theta)\pi(\theta)d\theta} \quad (2.3)$$

Na expressão (2.3), $f(\underline{x}/\theta)$ é a função de verossimilhança dado que é a distribuição condicional de \underline{X} aplicada no ponto \underline{x} , visto que a amostra \underline{x} já foi observada. Ela exprime a informação do parâmetro contida na amostra \underline{x} ; $\pi(\theta)$ é a distribuição inicial, ou priori, do parâmetro, e indica o conhecimento acerca do parâmetro antes de observada a amostra \underline{x} ; $\pi(\theta/\underline{x})$ é a distribuição final, ou posteriori do parâmetro θ , e indica o conhecimento acerca do parâmetro depois de observada a amostra \underline{x} , considerando a informação a priori do parâmetro. Note que $\int f(\underline{x}/\theta)d\theta$ não depende de θ podendo ser interpretada apenas como uma constante normalizadora, necessária para garantir que o numerador de (2.3) venha a ser uma distribuição de probabilidade para θ . Assim, é muito comum encontrar na literatura a definição de posteriori como sendo proporcional ao produto verossimilhança X Priori, ou seja:

$$\pi(\theta/\underline{x}) \propto f(\underline{x}/\theta)\pi(\theta) \quad (2.4)$$

2.3. Prioris conjugadas

Em geral, a priori é representada por uma forma funcional paramétrica, cujos parâmetros devem ser especificados de acordo com o conhecimento que se tenha sobre θ antes da amostra ser observada. Estes parâmetros indexadores das distribuições a priori são chamados de hiperparâmetros para distingui-lo do parâmetro de interesse θ .

O uso de prioris conjugadas, em geral, facilita a análise, pois a obtenção da posteriori envolverá apenas uma mudança nos hiperparâmetros. A idéia é que quando a priori especificada é tal que, combinada com a verossimilhança do modelo, resulte em uma posteriori da mesma família da priori. Quando isso ocorre se diz que a priori utilizada é conjugada a modelo.

Como exemplo, considere Y_1, Y_2, \dots, Y_n uma amostra aleatória da distribuição de Poisson com parâmetro θ . Sua função de probabilidade conjunta é dada por:

$$p(y/\theta) = \frac{e^{-n\theta} \theta^{\sum_{i=1}^n y_i}}{\prod y_i!} \propto e^{-n\theta} \theta^{\sum_{i=1}^n y_i}, \theta > 0 \quad (2.5)$$

Se atribuirmos ao parâmetro θ da distribuição de Poisson uma priori com distribuição Gama de parâmetros positivos α e β , ou seja:

$$\pi(\theta) \propto \theta^{\alpha-1} e^{-\beta\theta}, \text{ onde } \alpha, \beta > 0 \text{ e } \theta > 0 \quad (2.6)$$

Teremos como densidade a posteriori:

$$\pi(\theta/y) \propto \theta^{\alpha + \sum_{i=1}^n y_i - 1} e^{-(\beta+n)\theta} \quad (2.7)$$

A expressão (2.7) corresponde ao núcleo de uma densidade $Gama\left(\alpha + \sum_{i=1}^n y_i, \beta + n\right)$. Então, temos que a priori Gama é uma priori conjugada para o parâmetro θ do modelo de Poisson.

Agora, considere Y_1, Y_2, \dots, Y_n uma amostra aleatória de uma distribuição $Normal(\mu, \sigma^2)$, onde μ e $\tau = \frac{1}{\sigma^2}$ são desconhecidos. O seguinte conjunto de prioris para μ e σ^2 é dito prioris conjugadas ao modelo normal:

$$\mu \sim Normal(\mu_0, \sigma_0^2) \quad (2.8)$$

$$\tau \sim Gama(\alpha, \beta), \alpha, \beta > 0, \tau = \sigma^{-2} \quad (2.9)$$

Isto porque, ao considerar priori normal para μ , a posteriori resultante para μ , ou seja, $\pi(\mu/x)$ também será da família normal. Raciocínio análogo pode ser feito para τ . Convém ressaltar que, parametrizar a normal em termos da precisão τ , ou seja, atribuir priori para τ ao invés de para σ^2 , é muito utilizado na literatura por facilitar as contas de posterioris em modelos mais complexos que envolvam a normal.

2.3.1. Priori não-informativas

Quando não se tem nenhum tipo de informação prévia ou quando a informação de que se dispõe sobre o parâmetro de interesse não é confiável, se utiliza o que se chama de distribuição a Priori não informativa, ou Priori de referência.

Dependendo dos valores especificados para os hiperparâmetros, uma priori conjugada pode ser não informativa. No caso da distribuição normal, por exemplo, pode-se caracterizar uma priori não-informativa se o valor estabelecido para o hiperparâmetro da variância for muito grande.

2.4. Resultados a Posteriori

Todas as inferências dos parâmetros desconhecidos do modelo podem e são obtidas da distribuição de probabilidade a posteriori. Assim, podemos dizer que o principal resultado da Análise Bayesiana é a derivação dessa distribuição de probabilidade.

Com a determinação da distribuição de probabilidade a posteriori, a qual expressa o atual conhecimento sobre o parâmetro analisado em termos probabilísticos, podemos obter estatísticas descritivas dessa distribuição de probabilidade, como, por exemplo, a média, a moda e a mediana, e utilizá-las como estimativas pontuais para θ . Um dos estimadores pontuais para θ mais utilizados é a *moda* da distribuição a posteriori, ou seja, o valor que maximiza a posteriori, idéia análoga a utilizada na obtenção de estimadores na Inferência Clássica, que buscam a maximização da função de verossimilhança. A moda da distribuição a posteriori de θ pode ser interpretada como o valor mais provável para θ .

Além de estimadores pontuais para θ se pode obter também estimativas por intervalo para o mesmo, também baseado na distribuição a posteriori de θ . Em Inferência Bayesiana, o intervalo obtido para uma estimativa é denominado Intervalo de Probabilidade a Posteriori para θ , também conhecido como Intervalo de Credibilidade.

Na literatura Bayesiana os dois tipos de intervalos de credibilidade mais citados são o Intervalo de credibilidade Central e o HDR (High Density Region). A seguir se apresentam as definições para estes dois tipos de intervalos de credibilidade.

Seja $\pi(\theta/x)$ a distribuição a posteriori para θ com. O intervalo (C_1, C_2) é chamado de intervalo de credibilidade Central para θ , com probabilidade a posteriori igual a $(1 - \alpha)$, se e somente se C_1 e C_2 são tais que:

$$\int_{-\infty}^{C_1} f(\theta | \underline{x}) d\theta = \int_{C_2}^{+\infty} f(\theta | \underline{x}) = \alpha / 2 \quad (2.10)$$

O Intervalo de Credibilidade HDR (High Density Region) para θ com probabilidade $(1 - \alpha)$ é definido como sendo a região a posteriori que:

- a) A densidade da posteriori em qualquer ponto pertencente a essa região é maior ou igual a densidade da posteriori em qualquer ponto fora da região;
- b) A amplitude da região é a menor possível;
- c) A probabilidade da posteriori nessa região é igual a menos $(1 - \alpha)$.

Ou seja, o HDR é o intervalo de credibilidade com probabilidade a posteriori igual a $(1 - \alpha)$ com menor amplitude dentre todos os intervalos possíveis.

Quando a distribuição a posteriori é uma distribuição unimodal e simétrica tem-se que o Intervalo de Credibilidade HDR é igual ao Intervalo de Credibilidade Central.

O Intervalo de Credibilidade HDR, por ser a região que contém os valores mais prováveis para o parâmetro desconhecido, é tido como o estimador Bayesiano por intervalo mais intuitivo. O Intervalo de Credibilidade HDR só pode ser determinado quando a posteriori é derivada analiticamente. Quando isso não acontece e é preciso recorrer a métodos computacionais, como simulação MCMC, por exemplo, utilizam-se os Intervalos de Credibilidade Central para os parâmetros desconhecidos do modelo.

2.5. Distribuição Preditiva a Posteriori

Considere que uma amostra aleatória X_1, X_2, \dots, X_n com função de probabilidade $f(x/\theta)$ tenha sido observada. Se o objetivo é prever o comportamento de uma realização futura dessa função de probabilidade $f(x/\theta)$, ou seja, determinar, por exemplo, o valor esperado de X_{n+1} , deve-se conhecer a distribuição de X_{n+1} . A suposição básica é que X_1, X_2, \dots, X_n são variáveis independentes e identicamente distribuídas. Então, a distribuição de X_{n+1} é representada por $f(x_{n+1}/\theta)$, onde θ é desconhecido. Depois de ter sido observada a amostra \underline{x} nossa incerteza sobre θ está representada pela posteriori $\pi(\theta/\underline{x})$. Desta forma, a distribuição preditiva de X_{n+1} é determinada por:

$$f(x_{n+1}/\underline{x}) = \int_{\Theta} f(x_{n+1}/\theta) f(\theta/\underline{x}) d\theta \quad (2.11)$$

A expressão dada por (2.11) descreve o comportamento de X_{n+1} dada a atual informação sobre θ . Essa função é chamada de Função Preditiva a Posteriori.

2.6. Métodos de Monte Carlo baseados em Cadeias de Markov (MCMC)

Como explicitado anteriormente, a metodologia da Estatística Bayesiana se apóia no teorema de Bayes. Muitos cientistas do século 19, como Laplace e Gauss, por exemplo, demonstraram algum interesse nas técnicas Bayesianas, no entanto foram ignorados pelos estatísticos do começo do século 20, os chamados clássicos ou frequentistas, sob o argumento que os métodos Bayesianos só poderiam resolver problemas simples, onde a distribuição a *posteriori* pudesse ser escrita analiticamente. Para a obtenção dessas posteriores geralmente se fazia necessário que a escolha das distribuições a *priori* se fizesse de forma conveniente a facilitar a obtenção de distribuições a posteriori e, por isto, nem sempre representantes do verdadeiro pensamento do analista. Além disto, problemas mais

complicados requerem modelos mais complexos e, assim, o cálculo de integrais também mais complicadas.

Atualmente, com a evolução das técnicas computacionais, a análise Bayesiana de dados pôde avançar e eliminar as restrições sobre *prioris* e modelos. Com isso, os métodos de integração de Monte Carlo, em especial *Markov Chain Monte Carlo* (MCMC), tornaram-se populares entre os Bayesianos. Estes métodos fornecem uma metodologia para a obtenção de *posterioris* quando se possui um vetor de parâmetros desconhecidos de dimensão muito alta. A idéia dos métodos MCMC é obter uma amostra da distribuição a *posteriori* através da geração de cadeias de Markov, os quais são definidos como processos estocásticos que descrevem trajetórias onde quantidades sucessivas são descritas probabilisticamente de acordo com o valor de seu predecessor imediato.

Em muitos casos, estes processos estocásticos tendem a um equilíbrio e as quantidades limites seguem uma distribuição probabilística que não varia mais. Essa distribuição é conhecida como distribuição estacionária, ou distribuição de equilíbrio, ou ainda, distribuição invariante.

As técnicas MCMC possibilitam simular a distribuição a *posteriori* através da sua incorporação como distribuição de equilíbrio em uma cadeia de Markov. A idéia então, é simular esta cadeia até que ela atinja seu equilíbrio. Portanto, neste estado, os valores da cadeia de Markov podem ser vistos como uma amostra de valores (correlacionados) da distribuição a *posteriori* conjunta dos parâmetros. De posse desta amostra, podem-se calcular médias, medianas, percentis, probabilidades, enfim, qualquer característica desejada da amostra e utilizá-las como estimativas da *posteriori*. O algoritmo de Metropolis-Hastings , apresentado em 1970 e o Gibbs Sampler , de 1984, são os métodos mais utilizados para geração das cadeias de Markov.

A aplicação de técnicas MCMC exige algumas características específicas para a cadeia de Markov a ser simulada. Esta cadeia deve ser homogênea, irredutível, e

aperiódica. Para o leitor interessado, apresentam-se no anexo I estas definições e seus significados.

2.6.1. Algoritmo de Metropolis-Hastings

O algoritmo Metropolis-Hastings usa uma cadeia de Markov irreduzível e aperiódica e a partir desta constrói uma cadeia de Markov reversível no tempo cuja distribuição estacionária seja precisamente a distribuição a posteriori de interesse π .

O algoritmo de Metropolis-Hastings usa probabilidades de transição $Q = [q(i, j)]$ para construir uma cadeia de Markov $\{X_n, n \geq 0\}$, com matriz de transição $P = [p_{ij}]$, que é reversível no tempo e que tem a distribuição estacionária pretendida, no caso π . As probabilidades de transição $Q = [q(i, j)]$ são representadas por uma distribuição de probabilidade $q(j/i)$ chamada de distribuição proposta. A distribuição $q(j/i)$ tem por objetivo simular candidatos para o estado J da cadeia. Note que para configurar uma Cadeia de Markov esta distribuição depende do passado da cadeia somente através do estado i anterior.

A idéia do algoritmo é, então, gerar um ponto candidato θ , a cada tempo n , a partir da distribuição proposta $q(i/j)$ e aceitar esse ponto candidato como próximo estado da cadeia com uma probabilidade definida $\alpha(i, j)$. Essa idéia pode ser esboçada como se segue.

Se $\theta_n = i$, gera-se um valor θ da distribuição $q(i/j)$. Se $\theta = j$, faz-se $\theta_{n+1} = j$ com probabilidade $\alpha(i, j)$ e $\theta_{n+1} = i$ com probabilidade $1 - \alpha(i, j)$. Se o ponto candidato não for aceito, a cadeia não se move, isto é, $\theta_{n+1} = i$.

Deste modo, geramos uma cadeia de Markov com probabilidades de transição $P_{i,j}$ definidas por $P_{i,j} = q(i,j)\alpha(i,j)$, se $j \neq i$, e $P_{i,i} = q(i,i) + \sum_{k \neq i} q(i,k)(1 - \alpha(i,k))$

Esta cadeia de Markov é reversível no tempo se $\pi(i)P_{ij} = \pi(j)P_{ji}$ para $j \neq i$, o que é equivalente a:

$$\pi(i)q(i,j)\alpha(i,j) = \pi(j)q(j,i)\alpha(j,i). \quad (2.12)$$

Essa condição é verdadeira se fizermos

$$\alpha(i,j) = \min\left(\frac{\pi(j)q(j,i)}{\pi(i)q(i,j)}, 1\right) \quad (2.13)$$

Assim, podemos definir o algoritmo de Metropolis-Hastings da seguinte forma:

1. Considere uma matriz de transição de uma cadeia de Markov irredutível, $Q = [q(i,j)]_{i,j=1,2,\dots,m}$. Escolha um estado qualquer k , entre 1 e m .
2. Faça $n = 0$ e $\theta_n = k$. O valor k é o valor inicial e pode ser escolhido arbitrariamente.
3. Gere um valor θ da distribuição proposta $q(\theta_n, \cdot)$. Gere também um número aleatório U , de uma distribuição *Uniforme* $(0,1)$.
4. Se $U < \alpha(i,j)$, então $\theta_{n+1} = \theta$, caso contrário $\theta_{n+1} = \theta_n$.

5. Faça $n = n + 1$ e $\theta_n = \theta_{n+1}$.
6. Volte ao passo 3, até obter convergência.

2.6.2. Algoritmo Gibbs Sampling

O amostrador de Gibbs pode ser definido como um caso especial do algoritmo de Metropolis-Hastings. A principal diferença é que no método de Gibbs a cadeia irá sempre se mover para um novo valor, ou seja, não existe mecanismo de aceitação ou rejeição de um valor gerado.

Seja $\theta = (\theta_1, \theta_2, \dots, \theta_n)$ um vetor aleatório com dimensão n e com função densidade de probabilidade $\pi(\theta)$. Tem-se como objetivo gerar um vetor aleatório com essa distribuição de probabilidade. Para isso, usam-se as funções densidade condicionais completas de $\pi(\theta)$ com distribuições propostas, através do algoritmo abaixo. Note que todas as distribuições condicionais completas precisam ser conhecidas.

1. Fixe um valor inicial qualquer para a cadeia, $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_n^{(0)})$.
2. Faça $i=1$.
3. Obtenha o valor do estado através das seguintes gerações sucessivas:

$$\begin{aligned}
\theta_1^{(i)} &\sim \pi(\theta_1 / \theta_2^{(i-1)}, \theta_3^{(i-1)}, \dots, \theta_n^{(i-1)}) \\
\theta_2^{(i)} &\sim \pi(\theta_2 / \theta_1^{(i)}, \theta_3^{(i-1)}, \dots, \theta_n^{(i-1)}) \\
&\vdots \\
&\vdots \\
&\vdots \\
\theta_n^{(i)} &\sim \pi(\theta_n / \theta_1^{(i)}, \theta_2^{(i)}, \dots, \theta_{n-1}^{(i)})
\end{aligned}
\tag{2.14}$$

4. Faça $i=i+1$.

5. Volte ao passo dois até obter convergência.

2.6.3. A amostra da Posteriori

A amostra a posteriori obtida pelo Método de Gibbs Sampling não é uma amostra independente. Isso se deve ao fato de que o método de Gibbs se utiliza de Cadeias de Markov para a obtenção dessa amostra. Ou seja, cada valor gerado da amostra depende dos valores gerados na interação anterior. Assim, uma das maneiras mais usuais apresentadas na literatura para a obtenção de uma amostra aproximadamente independente baseia-se no fato que, após a convergência, os valores simulados seguem a distribuição estacionária $\pi(\theta)$. Assim, os b primeiros valores gerados são considerados como um período de aquecimento do algoritmo (da expressão em inglês “burn-in”), necessário para a convergência do Gibbs Sampling. Então estes b primeiros valores são descartados, considerando-se apenas os n valores posteriores como a amostra da distribuição de interesse. Mesmo assim, quando a dependência dos valores simulados é muito alta, pode-se determinar um valor k , conhecido como “thin” em inglês, pelo fato deste procedimento diminuir o tamanho da amostra. A idéia é que a amostra será formada apenas pelos valores gerados da cadeia a cada k -ésima iteração após o “burn-in”, ou seja, $\theta_i^{(b+1)}, \theta_i^{(b+k+1)}, \theta_i^{(b+2k+1)}, \dots$.

A determinação do período de “burn-in” não pode ser realizada analiticamente na maioria dos casos. Nestes casos, existem alguns métodos informais para estabelecer o tamanho deste período. O método mais utilizado baseia-se em gráficos da trajetória da cadeia, para verificar a partir de qual iteração os dados passam a apresentar o mesmo padrão de comportamento.

Considerando-se uma simulação por Gibbs Sampling através do algoritmo descrito na Subseção 2.6.2, obtêm-se uma amostra da qual se podem estimar todas as estatísticas de interesse, como média, variância, moda e mediana, por exemplo. Também através dessa amostra, pode-se estimar não-parametricamente a distribuição a posteriori, a partir do método “Kernel Smothing”.

3. TEORIA DA CREDIBILIDADE

Em 1914, um grupo de atuários norte-americanos criou a base para os modelos de credibilidade utilizados até hoje, buscando solução para um dos maiores problemas da área de tarifação de seguros: a heterogeneidade existente, tanto dentro das carteiras como entre as carteiras. Essa idéia pode ser estendida também para casos em que se queira solucionar o problema de heterogeneidade dentro e entre apólices, ou dentro e entre classes de risco, e até dentro e entre seguradoras. Assim, a teoria da credibilidade foi projetada com o objetivo de encontrar estimadores que equilibrassem a experiência de uma apólice com a experiência da carteira desta apólice. Ou ainda, estimadores que equilibrassem a experiência de uma determinada classe de segurados, com o conjunto de todas as classes de segurados, ou grupo.

Desta forma, podemos dizer que a Teoria da Credibilidade é uma técnica de estimação para os parâmetros de interesse que se baseia na obtenção de uma média ponderada entre a experiência de uma determinada classe de segurados e a experiência da carteira a qual essa classe pertence, considerando a heterogeneidade existente tanto dentro das classes como entre elas. Essa ponderação se dá através de um fator, denominado como “Z”, o qual ficou conhecido como fator de credibilidade.

Dessa forma, surgiu a chamada Teoria da Credibilidade Americana, também denominada como Teoria da Credibilidade de Flutuação Limitada, cujas idéias foram utilizadas com sucesso por cerca de 50 anos. A preocupação principal dessa ramificação da Teoria da credibilidade era em relação ao tamanho da experiência de uma apólice para que o fator de credibilidade Z assumisse valor um e então apenas a experiência individual de cada classe fosse considerada na obtenção de estimativas de credibilidade para a apólice.

Em 1967, Bühlmann desenvolveu a fundamentação teórica a partir do artigo publicado por Bailey, de 1950, o qual havia definido um estimador linear nas observações

para a obtenção de estimativas para o prêmio de credibilidade, via método de mínimos quadrados. Assim surgia uma segunda ramificação da Teoria da Credibilidade, denominada como Teoria da Credibilidade de Maior Exatidão ou Teoria da Credibilidade Européia, (Pereira, 1998).

A partir desse período, se desenvolveram grande parte dos modelos clássicos de credibilidade utilizados até hoje: Modelo de Bühlmann (1967), Modelo de Bühlmann-Straub (1970), Modelo de regressão Hachemeister (1975), Modelo Hierárquico de Jewell (1975), entre outros.

Este trabalho tem como base a teoria da Credibilidade Européia, ou Teoria da Credibilidade de Maior Exatidão, cujas definições básicas são apresentadas na Seção 3.1. Na Seção 3.2 se introduz a discussão do tamanho da experiência necessária de uma classe para que apenas a estimativa individual dessa classe seja considerada na obtenção de uma estimativa de credibilidade para a mesma. Na Seção 3.3 é descrito o modelo de Bühlmann, por sua importância histórica, e, na Seção 3.4, é descrito o modelo de Bühlmann-Straub, o qual será utilizado no exemplo prático deste trabalho. Ambos os modelos são apresentados sob o enfoque Clássico e sob o enfoque Bayesiano. Na Seção 3.5 demonstra-se as idéias introduzidas na Seção 3.2, através de um exercício de simulação, baseado no modelo de Bühlmann, onde se utilizou o software estatístico R.

3.1. Teoria da Credibilidade de maior exatidão

Como dito anteriormente, praticamente todos os modelos de Credibilidade utilizados hoje se desenvolveram a partir das idéias definidas sob a concepção da Teoria da Credibilidade de maior exatidão, ou simplesmente Teoria da Credibilidade Européia. Estes modelos, de uma forma geral, buscam determinar um estimador de credibilidade que seja linear em relação aos dados e que minimize o erro quadrático desta estimação.

As definições básicas da Teoria da Credibilidade europeia são apresentadas em Pereira (1998), e reproduzidas a seguir.

Imagine uma carteira cuja j -ésima apólice, no i -ésimo tempo, tem parâmetro de risco θ_{ij} , para $i=1, 2, \dots, t$ e $j=1, 2, \dots, k$. De forma a simplificar os cálculos se supõe que os parâmetros sejam constantes no tempo, denotando-se então por θ_j , derivados de uma mesma função de probabilidade.

Define-se Y_{ij} como a variável de experiência de sinistro. Essa variável, como já mencionado acima, pode ser número de sinistros, taxas de risco, montante de indenizações pagas, etc.

Supõe-se ainda que:

- As apólices $j=1, 2, \dots, k$ são independentes e identicamente distribuídas;
- $Y_{1j}, Y_{2j}, \dots, Y_{tj}$ são independentes e identicamente distribuídos condicionalmente a θ_j , o que representa independência entre as apólices e homogeneidade no tempo.

Em seguida, se define o primeiro e o segundo momentos como:

$$M(\theta_j) = E(Y_{ij} / \theta_j) \quad (3.1)$$

$$\sigma_{\theta}^2 = Var(Y_{ij} / \theta_j) \quad (3.2)$$

Onde $M(\theta_j)$ é o parâmetro de interesse. Definem-se ainda, para i e j quaisquer:

$$E(Y_{ij}) = E(M(\theta_j)) = \mu \quad (3.3)$$

$$\text{Var}(Y_{ij}) = \sigma_\varepsilon^2 + \sigma_\theta^2 \quad (3.4)$$

Onde:

$\sigma_\varepsilon^2 = \text{Var}[E(Y_{ij} / \theta_j)] = \text{Var}[M(\theta_j)]$ é a dispersão/heterogeneidade entre as apólices;

$\sigma_\theta^2 = E[\text{Var}(Y_{ij} / \theta_j)] = E(\sigma_\theta^2)$ é a dispersão/heterogeneidade dentro das apólices.

O interesse é estimar $M(\theta_j) = E(Y_{ij} / \theta_j)$ através de uma função $g(y)$. Esta estimação se dará restringindo-se $g(y)$ a um conjunto de funções lineares de tal forma que se tenha

$$g(y) = C_0 + C_1 Y_1 + \dots + C_t Y_t \quad (3.5)$$

$$g(y) = C_0 + \sum_{i=1}^t C_i Y_i \quad (3.6)$$

Retirando-se o índice j para simplificar a notação, o erro quadrático médio será expresso por

$$\varepsilon(g) = E[(M(\theta) - g(y))^2] \quad (3.7)$$

Obtêm-se o erro quadrático mínimo sobre $g(y)$ definindo-se a primeira derivada em relação a C_r e igualando a zero, ou seja:

$$E\left[\left(M(\theta) - C_0 - \sum_{i=1}^t C_i Y_i\right)\right] = 0 \quad \text{para } r=0 \quad (3.8)$$

$$E\left[Y_r \left(M(\theta) - C_0 - \sum_{i=1}^t C_i Y_i\right)\right] = 0 \quad \text{para } r=1, 2, \dots, t \quad (3.9)$$

Multiplicando (3.8) por $E(Y_{ij})$ e substituindo em (3.9), para $r=1, 2, \dots, t$, se tem que:

$$C_0 + \mu \sum_{i=1}^t C_i = \mu \quad (3.10)$$

$$C_r \sigma_\theta^2 + \sigma_\varepsilon^2 \sum_{i=1}^t C_i = \sigma_\varepsilon^2 \quad (3.11)$$

Por simetria, têm-se que $C_1 = C_2 = \dots = C_t = C$, logo:

$$C_0 + tC\mu = \mu \quad (3.12)$$

$$C\sigma_\theta^2 + \sigma_\varepsilon^2 tC = \sigma_\varepsilon^2 \quad (3.13)$$

Então,

$$C = \frac{\sigma_\varepsilon^2}{\sigma_\theta^2 + t\sigma_\varepsilon^2} = \frac{Z}{t} \quad (3.14)$$

$$C_0 = (1 - Z)\mu \quad (3.15)$$

E finalmente, substituindo (3.14) e (3.15) em (3.6) e recolocando o índice j , chega-se a famosa fórmula de Credibilidade:

$$\hat{M}(\theta_j) = Z\bar{Y}_j + (1-Z)\mu \quad (3.16)$$

onde $\bar{Y} = \frac{\sum_{i=1}^t Y_i}{t}$ e $Z = \frac{t\sigma_\varepsilon^2}{\sigma_\theta^2 + t\sigma_\varepsilon^2}$

Então, para a obtenção da estimativa de Credibilidade para a classe j , representada por $\hat{M}(\theta_j)$, é necessário estimar μ , σ_θ^2 e σ_ε^2 . Os diversos modelos que embasaram a teoria da Credibilidade como, por exemplo, o Modelo de Bühlmann e sua extensão, o modelo de Bühlmann-Straub, os quais serão apresentados nas seções seguintes, se utilizaram do método de estimação apresentado acima, para a obtenção de estimadores para os parâmetros μ , σ_θ^2 e σ_ε^2 .

3.2. Tamanho da experiência

Entende-se por tamanho da experiência a quantidade de informação que se tem a respeito da variável Y em cada classe de risco dentro de uma apólice.

Para melhor entender as idéias que serão aqui discutidas, considere como exemplo uma apólice de seguro de vida em grupo com N segurados. Dentro dessa apólice, os segurados foram classificados em k classes de risco, de acordo com características individuais de cada segurado. Observe ainda que a estimativa de credibilidade para cada classe de risco j será uma ponderação entre a média de Y no grupo de todos os N segurados pertencentes a esta apólice (representada em (3.16) por μ) e a média de Y da respectiva

classe (representada em (3.16) por \bar{Y}). Logo, se Z vale zero, diz-se que a credibilidade individual de cada classe é nula e somente a estimativa coletiva é considerada na obtenção da estimativa de credibilidade. No extremo contrário, tem-se Z igual a um, situação na qual está se atribuindo total credibilidade a classe e dispensando-se as informações da experiência do grupo

Baseado na fórmula da Teoria da Credibilidade, apresentada em (3.16), é natural pensar que, à medida que mais informações a respeito de uma classe são incorporadas, haverá um aumento no fator de credibilidade Z , o qual irá tender a um. Esta idéia é a base de todas as discussões em torno do chamado tamanho da experiência necessária para que Z seja um e uma classe tenha credibilidade completa na obtenção de uma estimativa de Credibilidade para a mesma. Essa discussão teve seu início com uma publicação de um artigo por Mowbray, no ano de 1914. (MANO, 1998). Neste artigo discutia-se o tamanho da experiência necessária para que o prêmio de uma determinada carteira pudesse ser estimado considerando-se unicamente a experiência da própria carteira. Segundo Mano, a discussão inicial proposta por Mowbray poderia ser reformulada para “quantos anos t de observação são necessários a fim de que a estimativa de credibilidade para uma determinada classe de risco possa ser obtida considerando-se apenas a média de Y dessa classe”.

A resposta para essa pergunta foi o principal objetivo da primeira ramificação da Teoria da Credibilidade, conhecida como Teoria da Credibilidade Americana. Segundo Pereira, (1998), grande parte dos estudos que procuravam encontrar uma resposta para essa pergunta se baseava numa aproximação pelo Teorema Central do Limite procurando determinar o tamanho da experiência necessária para que Z fosse igual a um, definindo dessa forma o que ficou denominado como Credibilidade Completa.

Na Seção 3.5 essa discussão será retomada, através de um exercício de simulação onde se mostrará a influência do tamanho da experiência de uma apólice na obtenção de estimativas de credibilidade para as classes de risco formadoras desta apólice.

3.3. Modelo de Bühlmann

Foi Bühlmann, em 1967, quem deu o primeiro formato a teoria da Credibilidade européia, apresentada na Seção 3.1.

O Modelo de Bühlmann é um modelo simples de credibilidade linear. Sua utilização prática se restringe a apólices balanceadas, ou seja, casos onde todas as classes de risco dentro de uma apólice possuem a mesma exposição. Entenda-se por exposição como sendo, por exemplo, o número de segurados expostos ao risco em cada classe.

Bühlmann desenvolve seu modelo a partir da hipótese de independência total das apólices, tanto dentro, supondo homogeneidade no tempo, quanto entre elas. Essas hipóteses, assim como as definições básicas para construção do modelo, já foram enunciadas inicialmente na Seção 3.2, e reproduzidas abaixo.

Imagine uma carteira cuja j -ésima apólice, no i -ésimo tempo, tem parâmetro de risco θ_{ij} , para $i=1, 2, \dots, t$ e $j=1, 2, \dots, k$. De forma a simplificar os cálculos se supõe que os parâmetros sejam constantes no tempo, denotando-se então por θ_j , derivados de uma mesma função de probabilidade.

Define-se Y_{ij} como a variável de experiência de sinistro. Essa variável, como já mencionado acima, pode ser número de sinistros, taxas de risco, montante de indenizações pagas, etc.

Supõe-se ainda que:

- As apólices $j=1, 2, \dots, k$ são independentes e identicamente distribuídas;

- $Y_{1j}, Y_{2j}, \dots, Y_{ij}$ são independentes e identicamente distribuídos condicionalmente a θ_j , o que representa independência entre as apólices e homogeneidade no tempo.

Supõe-se ainda o seguinte modelo:

$$y_{ij} = \theta_j + \varepsilon_{ij}, i = 1, 2, \dots, t, j = 1, 2, \dots, k \quad (3.17)$$

Onde:

$$\theta_j \sim Normal(\mu_\theta, \sigma_\theta^2) \quad (3.18)$$

$$\varepsilon_{ij} \sim Normal(0, \sigma_\varepsilon^2) \quad (3.19)$$

A partir das definições básicas, o modelo de Bühlmann se desenha sobre a seguinte matriz de dados, apresentada na Tabela 3.1.

Tempo	Parâmetros de risco			
	θ_1	θ_2	...	θ_k
1	Y_{11}	Y_{12}	...	Y_{1k}
2	Y_{21}	Y_{22}	...	Y_{2k}
...
t	Y_{t1}	Y_{t2}	...	Y_{tk}

Tabela 3.1 – estrutura de dados para aplicação do Modelo de Bühlmann

3.3.1 Abordagem Clássica para o Modelo de Bühlmann

Sob hipótese de normalidade dos dados os estimadores derivados por Bühlmann se reportam a um modelo extremamente difundido e muito utilizado na área da Estatística, conhecido como modelo de Componentes de Variância. Os estimadores apresentados por

Bühlmann, com Z definido como $Z = \frac{t\sigma_\varepsilon^2}{\sigma_\theta^2 + t\sigma_\varepsilon^2}$, são os seguintes:

$$\hat{\mu} = \frac{\sum_{j=1}^k \bar{Y}_j}{k} \quad (3.20)$$

$$\hat{\sigma}_\theta^2 = \frac{1}{k(t-1)} \sum_{j=1}^k \sum_{i=1}^t (Y_{ij} - \bar{Y}_j)^2 \quad (3.21)$$

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{k-1} \sum (\hat{\mu} - \bar{Y}_j)^2 - \frac{\hat{\sigma}_\theta^2}{t} \quad (3.22)$$

Os estimadores apresentados acima foram derivados via método de máxima verossimilhança. A idéia é encontrar os estimadores para os parâmetros μ , σ_θ^2 e σ_ε^2 e substituí-los na fórmula de Z , apresentada em (3.16)

3.3.2 Abordagem Bayesiana para o Modelo de Bühlmann

O modelo de Componentes de Variância apresentado anteriormente pode ser entendido em sua estrutura hierárquica da seguinte forma:

$$y_{ij} / \theta_j \sim \text{Normal}(\theta_j, \sigma_\varepsilon^2) \quad (3.23)$$

$$\theta_j \sim \text{Normal}(\mu_\theta, \sigma_\theta^2) \quad (3.24)$$

Sob a abordagem Bayesiana a distribuição estabelecida para θ é chamada de priori e seus parâmetros, (neste caso μ_θ e σ_θ^2), de hiperparâmetros. Para especificar a construção de um modelo completamente Bayesiano se faz necessário a especificação de distribuição a priori para os hiperparâmetros. Essas prioris são chamadas de hiperprioris e os valores de seus parâmetros devem ser escolhidos de forma a refletir o conhecimento a priori do pesquisador sobre eles. Ou seja, é através da especificação das hiperprioris que informações sobre os parâmetros desconhecidos que são alheios aos dados são diretamente incorporados na análise.

Portanto, para completar o modelo hierárquico supõe-se independência a priori dos hiperparâmetros, prioris conjugadas e parametrização da normal em termos de precisão. Mais especificamente, o modelo completo será:

$$y_{ij} / \theta_j \sim \text{Normal}(\theta_j, \sigma_\varepsilon^2) \quad (3.25)$$

$$\theta_j \sim \text{Normal}(\mu_\theta, \sigma_\theta^2) \quad (3.26)$$

$$\mu \sim \text{Normal}(\mu_0, \sigma_0^2) \quad (3.27)$$

$$(3.28)$$

$$\tau_\varepsilon = 1/\sigma_\varepsilon^2 \sim \text{Gama}(\alpha_1, \beta_1), \alpha, \beta > 0$$

$$\tau_\theta = 1/\sigma_\theta^2 \sim \text{Gama}(\alpha_2, \beta_2), \alpha, \beta > 0 \quad (3.29)$$

A inferência sob a abordagem Bayesiana depende da derivação da distribuição a posteriori para os parâmetros desconhecidos do modelo. Assim, neste modelo se faz necessário a derivação da posteriori conjunta $\pi(\theta_1, \dots, \theta_k, \mu_\theta, \sigma_\theta^2, \sigma_\varepsilon^2 / x)$. No entanto, esta distribuição conjunta não é derivável analiticamente. Porém todas as distribuições condicionais completas dos parâmetros são conhecidas. Assim, pode-se utilizar o método Gibbs Sampling para a derivação da posteriori conjunta $\pi(\theta_1, \dots, \theta_k, \mu_\theta, \sigma_\theta^2, \sigma_\varepsilon^2 / x)$.

As distribuições condicionais completas do modelo de componentes de variância, derivadas a partir de sua formulação hierárquica, com as distribuições a priori especificadas anteriormente são as seguintes:

$$\sigma_\theta^2 / y, \mu, \theta, \sigma_\varepsilon^2 \sim \text{IG}\left(\alpha_1 + \frac{1}{2}k; \beta_1 + \frac{1}{2}\sum_{j=1}^k (\theta_j - \mu)^2\right) \quad (3.30)$$

$$\sigma_\varepsilon^2 / y, \mu, \theta, \sigma_\theta^2 \sim \text{IG}\left(\alpha_2 + \frac{1}{2}kt; \beta_2 + \frac{1}{2}\sum_{i=1}^t \sum_{j=1}^k (y_{ij} - \theta_j)^2\right) \quad (3.31)$$

$$\mu / y, \theta, \sigma_\theta^2, \sigma_\varepsilon^2 \sim N\left(\frac{\sigma_\theta^2 \mu + \sigma_0^2 \sum_{j=1}^k \theta_j}{\sigma_\theta^2 + k\sigma_0^2}, \frac{\sigma_\theta^2 \sigma_0^2}{\sigma_\theta^2 + k\sigma_0^2}\right) \quad (3.32)$$

$$\theta_j / y, \mu, \sigma_\theta^2, \sigma_\varepsilon^2 \sim N\left(\frac{t\sigma_\varepsilon^2}{t\sigma_\varepsilon^2 + \sigma_\theta^2} \bar{y}_{\cdot j} + \frac{\sigma_\varepsilon^2}{t\sigma_\varepsilon^2 + \sigma_\theta^2} \mu; \frac{\sigma_\theta^2 \sigma_\varepsilon^2}{t\sigma_\varepsilon^2 + \sigma_\theta^2}\right) \quad (3.33)$$

Onde $\bar{y}_{.j} = \frac{1}{k} \sum_{h=1}^k y_{jh}$ e IG se refere a distribuição Inversa-Gama.

Note que o parâmetro que representa a média da distribuição condicional completa de θ é exatamente a fórmula de Credibilidade, apresentada em (3.16)

3.4. Modelo de Bühlmann-Straub

O chamado modelo de Bühlmann-Straub, desenvolvido em 1970, é uma extensão daquele definido por Bühlman, com objetivo de solucionar o problema de quando a carteira tem exposições diversas em cada apólice, ou seja, o número de segurados de cada apólice não é o mesmo em todas as apólices da carteira. A única variação que se dá neste modelo em relação ao modelo de Bühlmann é a inclusão de pesos para ponderar a importância de cada apólice dentro de uma carteira.

3.4.1. Abordagem Clássica para o Modelo de Bühlmann-Straub

Definindo-se w_{ij} como os pesos de cada apólice j no tempo i , tem-se o estimador de credibilidade definido por Bühlmann-Straub dado por:

$$M_j = Z_j Y_{jw} + (1 - Z_j) \mu \quad (3.34)$$

Sendo $Z_j = \frac{\sigma_\varepsilon^2 w_{\cdot j}}{\sigma_\varepsilon^2 w_{\cdot j} + \sigma_\theta^2}$ e os estimadores para $\mu, \sigma_\theta^2, \sigma_\varepsilon^2$, obtidos pelo método da

máxima verossimilhança, os seguintes:

$$\hat{\mu} = \sum_{j=1}^k \frac{Z_j}{Z_{\cdot}} Y_{jw} = \sum_{i=1}^t \sum_{j=1}^k \frac{Z_j}{Z_{\cdot}} \frac{w_{ij}}{w_{\cdot j}} Y_{ij} \quad (3.35)$$

$$\sigma_\theta^2 = \frac{1}{k(t-1)} \sum_{i=1}^t \sum_{j=1}^k w_{ij} (Y_{ij} - Y_{jw})^2 \quad (3.36)$$

$$\sigma_\varepsilon^2 = \frac{w_{\cdot\cdot}}{w_{\cdot\cdot}^2 - \sum_{j=1}^k w_{\cdot j}^2} \left[\sum_{j=1}^k w_{\cdot j} (Y_{jw} - Y_{ww})^2 - (k-1) \sigma_\theta^2 \right] \quad (3.37)$$

Onde as seguintes notações são utilizadas:

$$w_{\cdot\cdot} = \sum_{j=1}^k w_{\cdot j} = \sum_{i=1}^t \sum_{j=1}^k w_{ij} \quad (3.38)$$

$$Z_{\cdot} = \sum_{j=1}^k Z_j \quad (3.39)$$

$$Y_{jw} = \sum_{i=1}^t \frac{w_{ij}}{w_{\cdot j}} Y_{ij} \quad (3.40)$$

$$Y_{ww} = \sum \frac{w_{\cdot j}}{w_{\cdot\cdot}} Y_{jw} = \sum_{i=1}^t \sum_{j=1}^k \frac{w_{ij}}{w_{\cdot\cdot}} Y_{ij} \quad (3.41)$$

3.4.2. Abordagem Bayesiana para o Modelo de Bühlmann-Straub

Em sua forma hierárquica o modelo de Bühlmann-Straub se apresenta da seguinte forma:

$$y_{ij} / \theta_j \sim Normal(\theta_j, \sigma_\varepsilon^2) \quad (3.42)$$

$$\theta_j \sim Normal\left(\mu_\theta, \frac{\sigma_\theta^2}{w_{ij}}\right) \quad (3.43)$$

Assim, a única alteração do modelo de Bühlmann-Straub em relação ao modelo de Bühlmann é sobre a variância, a qual passa a ser $Var(Y_{ij} / \theta_j) = \frac{\sigma_\theta^2}{w_{ij}}$. Desta forma, todas as especificações necessárias para completar o modelo, a saber, as prioris para os hiperparâmetros são definidas como na seção anterior. Consequentemente, os procedimentos inferenciais são também obtidos através do uso de Gibbs Sampling, onde as condicionais completas são análogas as já apresentadas para o modelo de Bühlmann.

3.5. Simulação aplicada ao Modelo de Bühlmann

Esta seção tem por objetivo mostrar as idéias sobre o tamanho da experiência apresentadas na Seção 3.2 aplicadas ao Modelo de Bühlmann, através de um exercício de simulação, ou seja, pretende-se demonstrar que, à medida que aumenta o histórico de observações de cada classe, maior será a credibilidade dessa classe na obtenção de estimativas de credibilidade para a mesma. Todas as simulações apresentadas foram realizadas no software R e a programação utilizada está apresentada no Anexo 2.

Para iniciar foram simulados conjuntos de dados que representassem o modelo de Bühlmann apresentado na Seção 3.3 de acordo com os seguintes valores para os parâmetros: número de classe $k=5$, $\mu = 100$, $\sigma_{\theta}^2 = 25$, $\sigma_{\epsilon}^2 = 10$. Definiram-se também os seguintes tamanhos de experiência $t = 2, 3, 5, 8$ e 12 anos. Este processo foi repetido 100 vezes, ou seja, gerou-se 100 conjuntos de dados para cada tamanho de experiência t , de acordo com a seguinte estrutura:

Tempo	Parâmetros de risco			
	θ_1	θ_2	...	θ_k
1	Y_{11}	Y_{12}	...	Y_{1k}
2	Y_{21}	Y_{22}	...	Y_{2k}
...
t	Y_{t1}	Y_{t2}	...	Y_{tk}

Tabela 3.2 – Estrutura de dados para o modelo de Bühlmann

Para cada conjunto de dados simulados obtiveram-se, através do método de Gibbs Sampling, estimativas para todos os parâmetros do modelo.

Adotaram-se distribuições a priori não informativas para todos os parâmetros do modelo. Assim, os valores definidos para os hiperparâmetros, parâmetros das distribuições a priori, foram os seguintes: $\mu = 100$, $\sigma_{\theta}^2 = 10^4$, $\alpha_1 = 0$, $\beta_1 = 0$, $\alpha_2 = 0$, $\beta_2 = 0$. Simulou-se uma cadeia de tamanho $m=10000$ iterações, com um período de “burn-in” de $b=3000$ iterações. Para cada cadeia gerada, desconsiderava-se o período de burn-in na obtenção das estimativas para os parâmetros do modelo.

Para a demonstração objetivada neste exemplo, foram observados apenas os valores estimados de θ_j . Então, tomou-se, para cada classe, o quadrado da diferença entre a

estimativa obtida e a média da respectiva classe, ou seja: $d1 = (\theta_j - \bar{Y}_j)^2$. Também se tomou o quadrado da diferença entre a estimativa de cada classe e a média geral de todas as classes: $d2 = (\theta_j - \bar{Y})^2$. À medida que Z se aproxima de um, maior é o peso da média da classe j na estimativa θ_j , ou seja, menor é a diferença entre a estimativa θ_j e a média da classe j , representada por \bar{Y}_j . Assim, na medida em que o tempo passa e mais informações são incorporadas, a estimativa θ_j tende a ser a própria média \bar{Y}_j e por consequência a diferença $(\theta_j - \bar{Y}_j)$ tende a zero. Consequentemente, à medida que a estimativa θ_j se aproxima de \bar{Y}_j a mesma se distancia de \bar{Y} . Logo, a diferença $(\theta_j - \bar{Y})$ tende a crescer à medida que o tempo passa e mais informações são incorporadas.

Resultados

Na Figura 3.1 estão os gráficos Box-Plot para o quadrado da diferença entre a estimativa θ_1 e a média da classe de risco I, representada por \bar{Y}_1 , e para o quadrado da diferença entre a estimativa θ_1 e a média de toda a apólice, representada por \bar{Y} . Pode-se perceber através destes gráficos que, quanto maior o tamanho da experiência da carteira menor será a diferença entre a estimativa θ_1 e a média da classe um, significando que, com o correr do tempo, apenas a estimativa individual da classe um pode ser considerada na obtenção de estimativas de credibilidade para a mesma. Por outro lado, percebe-se um aumento da diferença entre θ_1 e a média coletiva de todas as classes de risco da apólice.

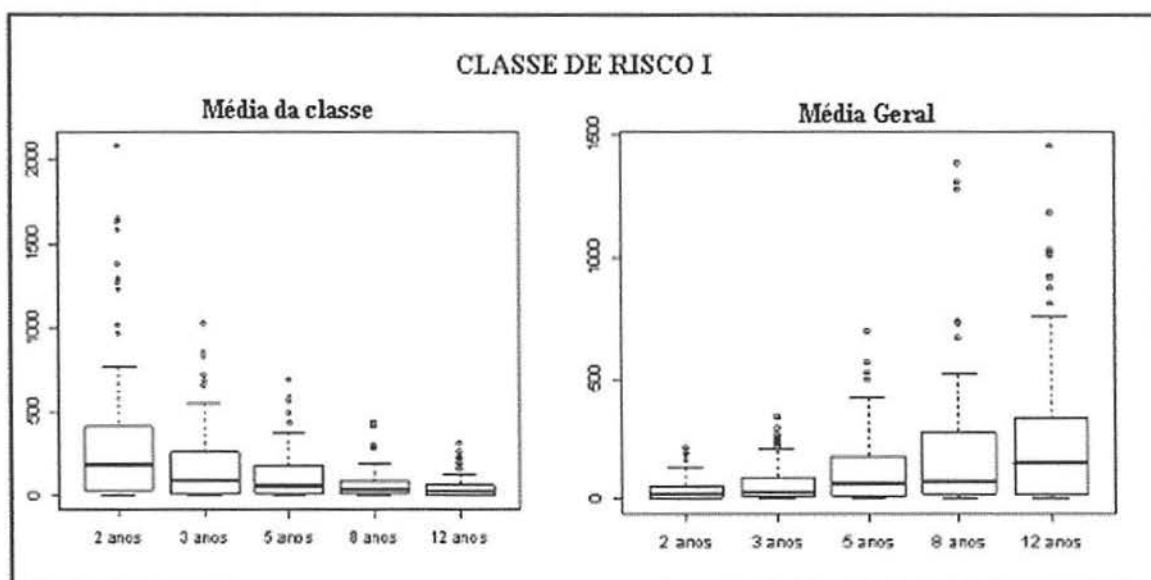


Figura 3.1 – Box-Plot das diferenças $(\theta_1 - \bar{y}_1)^2$ e $(\theta_1 - \bar{y})^2$, por tamanho de experiência

Analogamente, observa-se a mesma situação para as demais classes de risco assumidas neste exercício, conforme apresentado nas figuras 3.2, 3.3, 3.4 e 3.5.

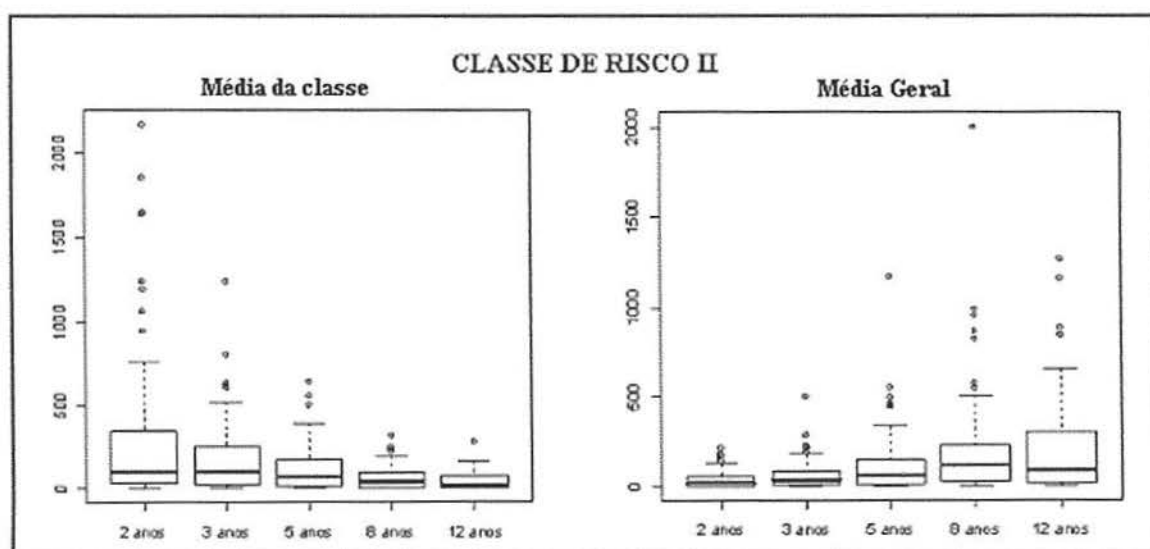


Figura 3.2 – Box-Plot das diferenças $(\theta_2 - \bar{Y}_2)^2$ e $(\theta_2 - \bar{Y})^2$, por tamanho de experiência

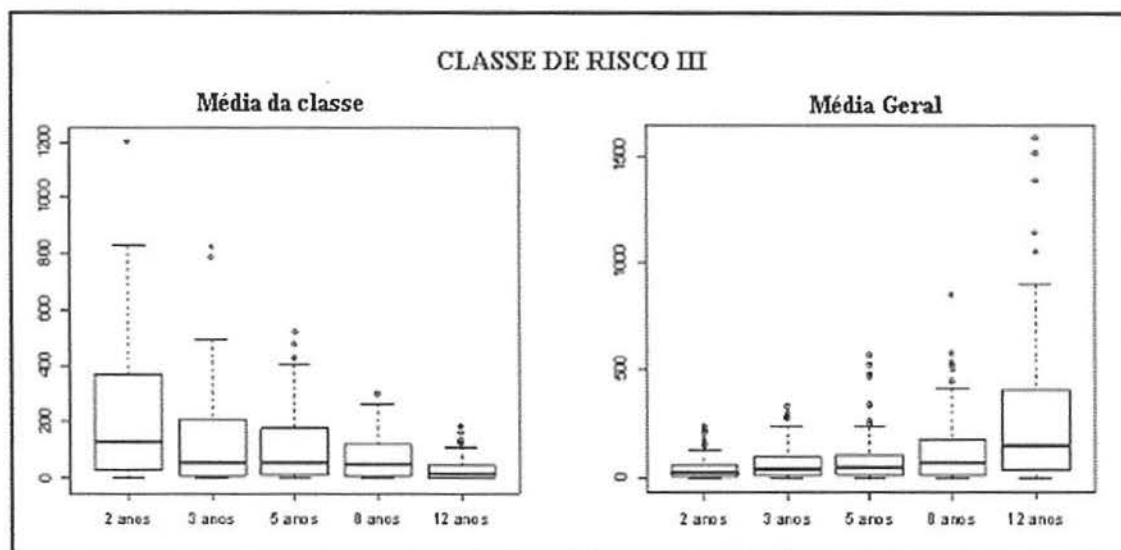


Figura 3.3 – Box-Plot das diferenças $(\theta_3 - \bar{Y}_3)^2$ e $(\theta_3 - \bar{Y})^2$, por tamanho de experiência

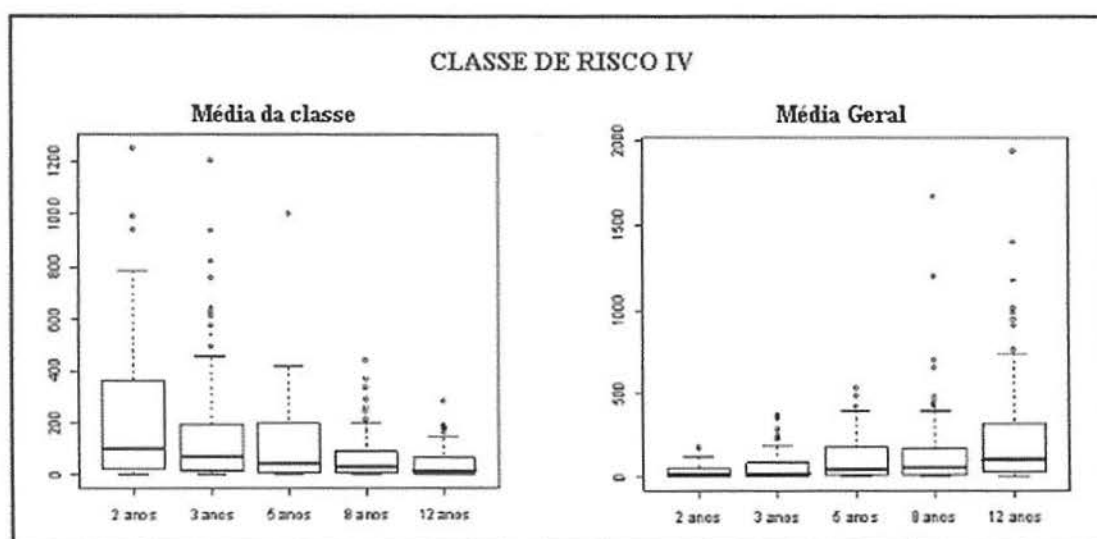


Figura 3.4 – Box-Plot das diferenças $(\theta_4 - \bar{Y}_4)^2$ e $(\theta_4 - \bar{Y})^2$, por tamanho de experiência

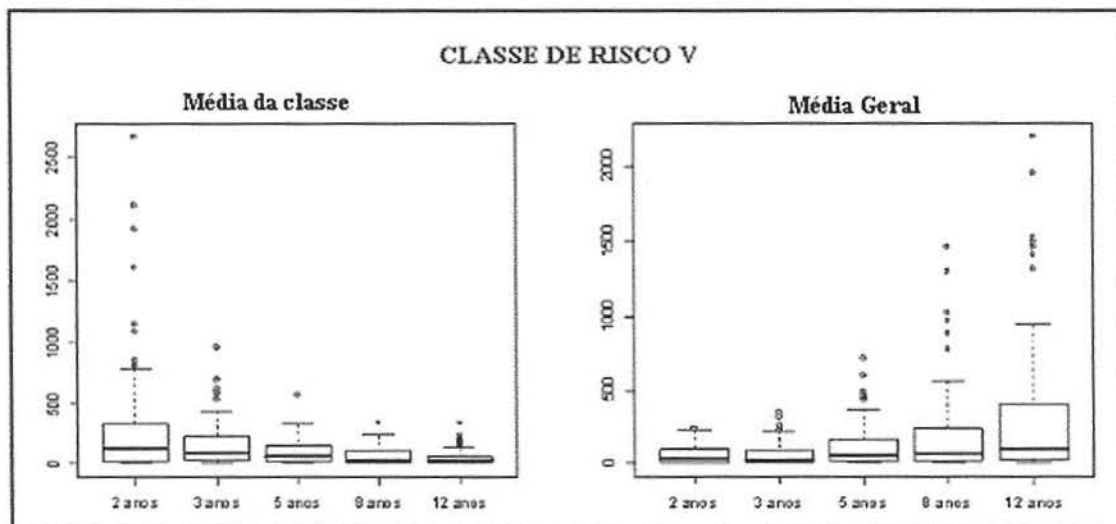


Figura 3.5 – Box-Plot das diferenças $(\theta_s - \bar{Y}_s)^2$ e $(\theta_s - \bar{Y})^2$, por tamanho de experiência

4. EXEMPLO PRÁTICO APLICADO EM SEGUROS DE VIDA EM GRUPO

Este capítulo tem como objetivo ilustrar com um exemplo prático a aplicação da Teoria da Credibilidade na obtenção de estimativas de credibilidade para uma apólice de seguro de vida em grupo. Este seguro objetiva garantir uma indenização aos beneficiários do segurado, caso ocorra a morte do mesmo, independente da causa do óbito.

Os fundamentos atuariais utilizados como base técnica para o estudo da tarifação de seguros de vida podem ser resumidos nas chamadas Tábuas Biométricas, também denominadas Tábuas de Mortalidade ou Tábuas de Sobrevivência, as quais representam a experiência para a população em estudo. A partir de uma determinada tábua de mortalidade, específica para cada tipo ou situação de cálculo, se obtém, por exemplo, o número esperado de sinistros para um período futuro, período este geralmente definido em um ano.

As normas para adoção de tábuas biométricas costumam indicar tábuas de experiências internacionais como adequadas, de acordo com a situação de cálculo. Atualmente, as Tábuas de Mortalidade admitidas no Brasil para o cálculo dos prêmios do Seguro Vida em Grupo, são: SGB-71, CSO-58 MALE, CSO-80 MALE, SCSG-60, GKM-70 MALE, ALLG-72 MALE E AT-83 MALE, podendo ser utilizadas outras tábuas, desde que reconhecidas pelo Instituto Brasileiro de Atuária (IBA).

Definida como "o instrumento destinado a medir as probabilidades de vida e de morte", uma Tábua Biométrica consiste, na sua forma mais elementar, em uma tabela que registra, de um grupo inicial de pessoas da mesma idade, o número daqueles que vão atingindo as diferentes idades, até a extinção completa do referido grupo. A Tábua de Mortalidade possui, na generalidade dos casos, quatro colunas com algarismos, sendo a primeira relativa às idades (x), a segunda ao número de sobreviventes, (l) x , a terceira ao

número de mortos $(d)_x$, e a quarta, e última, $(q)_x$ ao quociente da divisão de dx por lx , em cada linha. Uma ilustração de Tábua Biométrica é apresentada na Tabela 4.1.

x	l(x)	d(x)	q(x)
5	1.000.000	342	0,000342
6	999658	318	0,000318
7	999340	302	0,000302
8	999038	294	0,000294
9	998745	292	0,000292
10	998453	293	0,000293
11	998160	297	0,000298
•	•	•	•
•	•	•	•
•	•	•	•

Tabela 4.1 – Exemplo de Tábua Biométrica, GAM-83

Neste exemplo de aplicação se propõe ilustrar uma possível forma de aplicação da Teoria da Credibilidade, a qual consiste em prever o número de sinistros de uma carteira para um período futuro, atribuindo-se um conjunto de prioris não-informativas (Solução I) e um conjunto de prioris informativas (Solução II) para os parâmetros desconhecidos do modelo. Em seguida se fará uma comparação entre as previsões de credibilidade obtidas e o número esperado de sinistros previsto pela tábua Biométrica AT-83, a qual é utilizada de forma mais freqüente na tarifação de prêmios em diversas seguradoras do Brasil. Os resultados serão comparados também com os valores reais observados para período cinco.

O objetivo desse exemplo é apenas ilustrar a aplicação da técnica da Teoria da Credibilidade sob um enfoque Bayesiano. Não há a preocupação em discutir as vantagens ou desvantagens da utilização de métodos Bayesianos em relação aos métodos Clássicos. Também não haverá preocupação em definir qual o melhor modelo a ser utilizado, entre os Modelos de Credibilidade. Isso devido ao fato de que, na prática, a escolha de um ou outro modelo depende muito da situação em que se está trabalhando e as variáveis que estão sendo utilizadas na tarifação do prêmio em questão. Também não será considerada ou

discutida a aceitabilidade da suposição de normalidade para a variável que será modelada, no caso, número de sinistros.

4.1. Descrição do banco de dados

Objetivando aplicação da Teoria da Credibilidade sob um enfoque Bayesiano, utilizaram-se dados reais da Seguradora Previdência do Sul. Foram utilizadas observações referentes a uma apólice de seguro de vida em grupo, em um período de cinco anos. As apólices foram divididas em cinco classes de acordo com a faixa etária dos segurados, denominadas de classes de risco. A obtenção dessas classes nesse exemplo foi definida de forma arbitrária, sendo irrelevante o conhecimento das mesmas. Foram utilizados na análise apenas os quatro primeiros períodos para que se possa comparar a previsão para o quinto período com o valor real observado, e também com o número esperado de sinistros pela Tábua Biométrica AT-83. As seguintes notações serão utilizadas neste exemplo prático:

Y_{ij} : número de sinistros ocorridos na classe de risco j no ano i ;

σ_{θ}^2 : variância dentro das classes;

σ_{ϵ}^2 : variância entre as classes;

$\tilde{Y}_{i+1,j}$: previsão de Y da classe de risco j para o período $i+1$.

A Tabela 4.2 e a Tabela 4.3 apresentam o número de segurados e de sinistros ocorridos para a apólice estudada:

Ano	Classes de Risco					Total
	I	II	III	IV	V	
1	4.272	5.403	5.132	3.398	845	19.050
2	4.098	5.289	4.964	3.217	781	18.349
3	3.902	5.143	4.779	3.034	699	17.557
4	3.731	4.917	4.557	2.819	620	16.644
Total	16.003	20.752	19.432	12.468	2.945	71.600

Tabela 4.2 - Número de segurados, por classe de risco e período

Ano	Classes de Risco					Total
	I	II	III	IV	V	
1	21	63	85	120	55	344
2	22	59	95	95	49	320
3	20	61	90	88	51	310
4	29	60	78	130	46	343
Total	92	243	348	433	201	1.317

Tabela 4.3 - Número de sinistros, por classe de risco e período

4.2.Procedimentos utilizados para análise

Como se pode observar nas tabelas anteriores, existe diferentes níveis de exposição ao risco em cada classe e em cada período, ou seja, o número de segurados não é o mesmo em todas as classes e também não se mantém constante ao passar do tempo. Dessa forma, se utilizou o modelo de Bühlmann-Straub, apresentado na Seção 3.4, o qual define pesos para os diferentes níveis de exposição, solucionando o problema de exposições diversas em cada classe.

As soluções foram obtidas através do método Gibbs Sampling, com simulação de cadeias com 10.000 iterações, com um período de burn-in de 3.000 iterações. Neste exemplo, a análise de convergência foi feita apenas através de inspeção visual, através dos gráficos das trajetórias das cadeias. No entanto, existem técnicas mais formais de análise de convergência em simulações pelo Método de Gibbs. Uma leitura mais completa sobre esse assunto pode ser encontrada em *Gilks et al (1997)*.

O programa utilizado neste exemplo foi o software WINBUGS. Este software é gratuito e pode ser obtido em: www.mrc-bsu.cam.ac.uk/bugs. A programação utilizada está apresentada no Anexo 3.

4.3. Análises e resultados

Solução I:

A solução proposta neste exemplo de aplicação, com o objetivo principal de prever o número de sinistros para o quinto ano da série, será de acordo com os seguintes passos:

Passo 1: definição de um conjunto de prioris não informativas para μ , σ_θ^2 e σ_ε^2 .

Naturalmente, se utilizou de prioris conjugadas para o modelo normal, então:

$$\begin{aligned}\mu &\sim Normal(0, 1.0E^{-5}) \\ 1/\sigma_\theta^2 &\sim Gamma(1.0E^{-3}, 1.0E^{-3}) \\ 1/\sigma_\varepsilon^2 &\sim Gamma(1.0E^{-3}, 1.0E^{-3})\end{aligned}$$

Passo 2: obter via amostrador de Gibbs a distribuição a posteriori e preditiva a posteriori para os parâmetros desconhecidos do modelo, de onde se obtêm as estimativas de interesse.

Resultados solução I:

A Figura 4.1 apresenta as trajetórias dos valores simulados para μ , θ_1 e \tilde{Y}_1 , desconsiderando-se o período de “burn-in”. Gráficos similares foram observados para todos os demais parâmetros do modelo, no entanto foram omitidos, apresentando-se apenas os gráficos para μ , θ_1 e \tilde{Y}_1 . Analisando estas trajetórias podemos considerar que parece haver convergência na simulação.

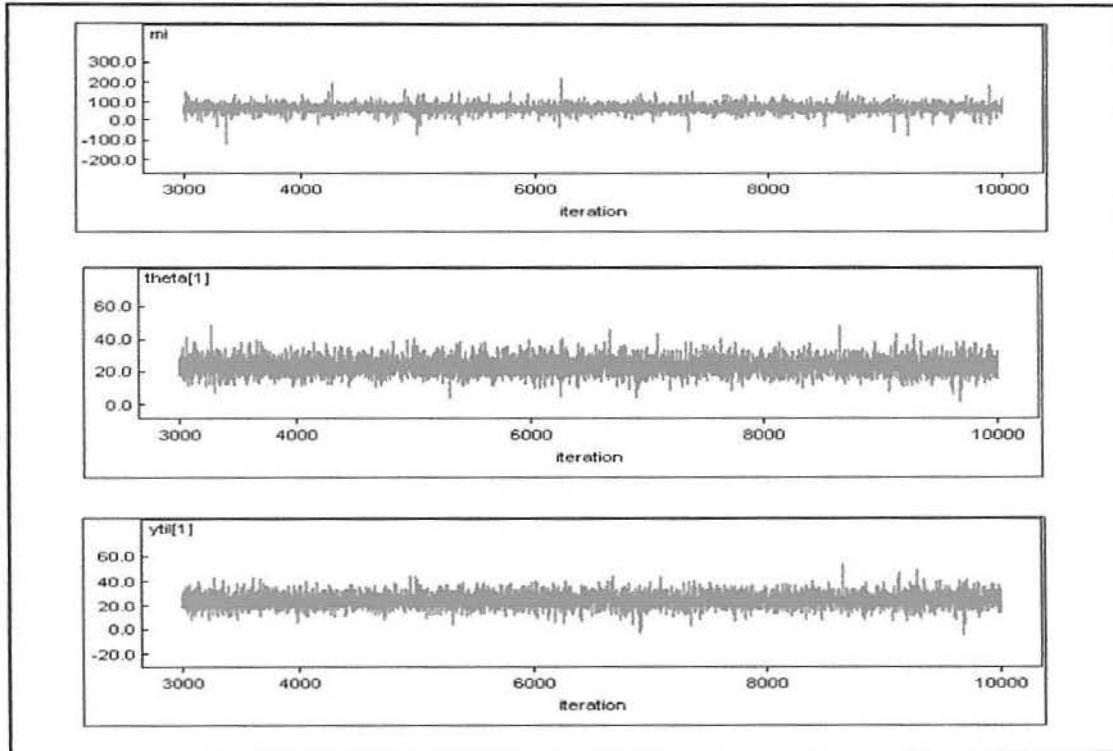


Figura 4.1 – Gráfico das trajetórias da simulação de μ , θ_1 e \tilde{Y}_1 , solução I

Os gráficos de densidade de Kernel apresentam estimativas não paramétricas das distribuições a posteriori. Estes gráficos podem ser observados na Figura 4.2, novamente para μ , θ_1 e \tilde{Y}_1 . Estas estimativas, obtidas pelo método de Kernel, representam o núcleo das respectivas distribuições a posteriori. Analisando os gráficos, pode-se perceber certa simetria nas distribuições a posteriori dos parâmetros. Com isso, a média a posteriori dos parâmetros é uma boa estimativa para os mesmos.

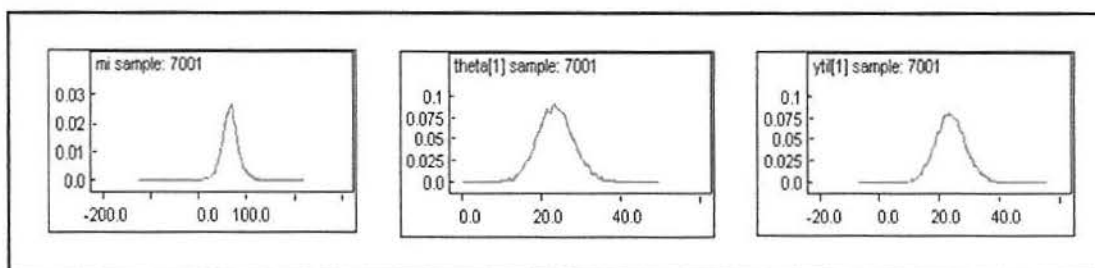


Figura 4.2 – Gráfico dos núcleos das densidades a posteriori de μ , θ_1 e \tilde{Y}_1 , solução I

A Tabela 4.4 apresenta os resultados descritivos a posteriori para os parâmetros do modelo. Os resultados de \tilde{Y} serão utilizados para prever o número de sinistros de cada classe de risco para período cinco. Esta previsão, obtida a partir de simulação pelo Método Gibbs Sampling, é comparada com o valor real observado e com o valor esperado pela Tábua Biométrica AT-83.

Parâmetros	média	Desv.Pad.	2.5%	97.5%	mediana
μ	65,830	19,800	26,980	105,000	65,970
σ_ε^2	0,0001	0,0007	0,0001	0,0002	0,0008
σ_θ^2	0,2334	0,0865	0,0977	0,4328	0,222
θ_1	23,900	4,776	14,800	33,600	23,650
θ_2	61,000	4,133	52,800	69,100	60,870
θ_3	87,000	4,259	78,300	91,200	86,710
θ_4	107,000	5,437	95,800	117,500	107,000
θ_5	52,000	10,380	31,400	72,600	52,000
\tilde{Y}_1	23,900	5,266	13,700	34,500	23,700
\tilde{Y}_2	61,000	4,676	51,500	70,300	61,000
\tilde{Y}_3	87,000	4,831	77,200	96,400	87,000
\tilde{Y}_4	107,000	5,831	95,000	118,200	107,000
\tilde{Y}_5	52,000	10,580	30,900	73,200	52,100

Tabela 4.4 – Resultados a posteriori para os parâmetros do modelo, solução I

A Tabela 4.5 apresenta um quadro comparativo entre a solução apontada acima com a estimativa obtida pela Tábua Biométrica AT-83 e o valor real observado para o período cinco.

Classe	Observado	AT-83	Solução I
I	22,0	14,0	23,9
II	58,0	19,0	61,0
III	86,0	35,0	87,0
IV	106,0	84,0	107,0
V	51,0	41,0	52,0
Total	323,0	193,0	330,9

Tabela 4.5 – Comparativo entre valores preditos, observados e esperados pela Tábua AT-83 para o Ano 5

Os resultados apresentados na Tabela 4.5 mostram que a Tábua Biométrica AT-83 fornece valores subestimados para o número de sinistros do ano 5, em todas as classes de risco. Por consequência, a previsão para o número total de sinistros também é subestimada. Enquanto espera-se, pela tábua, um número total de 193 sinistros para o ano 5, o observado para este período foi de 323 sinistros.

Por outro lado, as estimativas de credibilidade, obtidas utilizando-se prioris não informativas no modelo de Bühlmann-Straub apresentaram estimativas mais coerentes com o valor observado para o número de sinistros do ano 5. Na Figura 4.3 pode-se observar que os valores observados estão todos dentro do intervalo de credibilidade e muito próximas do valor previsto pontualmente para o número de sinistros do ano 5.

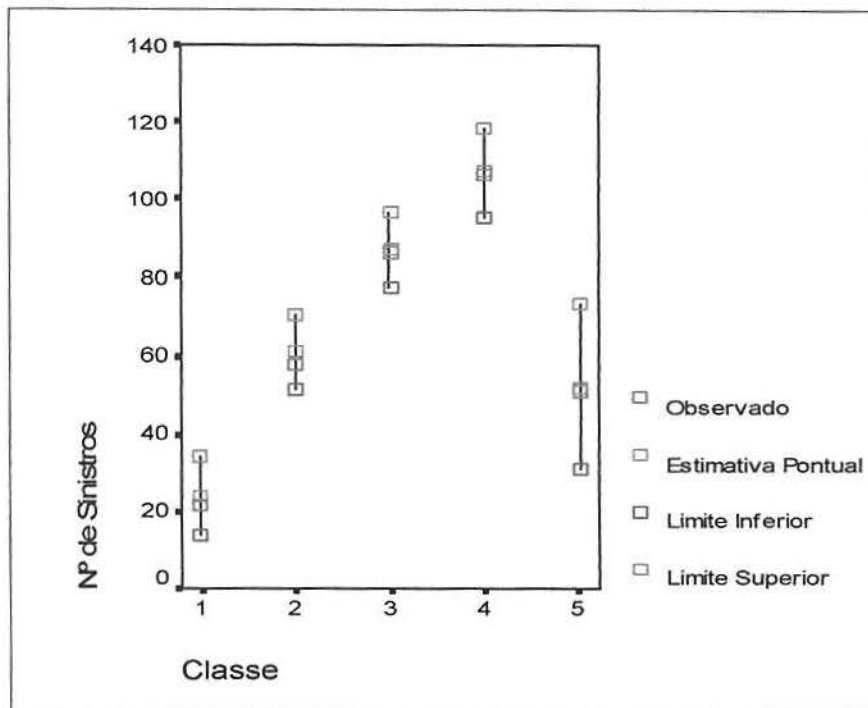


Figura 4.3 – Gráfico comparativo entre estimativas de credibilidade e valores observados

Solução II:

Apenas com fins ilustrativos, se propõe uma segunda solução utilizando o mesmo modelo utilizado na solução I, apenas alterando os valores fixados para as prioris de modo a torná-los mais informativos. Portanto, essa solução consiste em prever o número de sinistros de cada classe para o período cinco, combinando o número de sinistros esperados a partir de uma Tábua Biométrica com a experiência observada dessa carteira. As informações fornecidas pela Tábua Biométrica serão incorporadas no modelo através da atribuição de uma priori informativa para μ . Para os parâmetros σ_{θ}^2 e σ_{ε}^2 as prioris seguirão não informativas. Assim, a Solução 2 se deu a partir dos seguintes passos:

Passo 1: definição de um conjunto de prioris não informativas para μ , σ_θ^2 e σ_ε^2 . Naturalmente, se utilizou de prioris conjugadas para o modelo normal, então:

$$\begin{aligned}\mu &\sim \text{Normal}(39,10) \\ 1/\sigma_\theta^2 &\sim \text{Gamma}(1.0E^{-3},1.0E^{-3}) \\ 1/\sigma_\varepsilon^2 &\sim \text{Gamma}(1.0E^{-3},1.0E^{-3})\end{aligned}$$

Passo 2: obter via amostrador de Gibbs a distribuição a posteriori e preditiva a posteriori para os parâmetros desconhecidos do modelo, de onde se obtêm as estimativas de interesse.

Os valores fixados para a hiperpriori de μ têm por objetivo incorporar a informação não proveniente dos dados observados no modelo. O valor 39 representa a média de θ_j de acordo com a Tábua biométrica AT-83. Já o valor 10 representa a precisão da distribuição de μ . Note que, para a magnitude dos dados, este valor pode ser considerado pequeno. À medida que se queira dar mais ponderação a informação a priori de μ maior deve ser a precisão atribuída. Neste exemplo, não se dará importância a questão de como essa precisão deve ser atribuída e tão pouco do quanto a precisão atribuída a distribuição irá influenciar as estimativas obtidas. No entanto, faz-se necessário destacar que em uma aplicação prática, onde se queiram incorporar informações alheias aos dados no modelo para obter uma estimativa de Credibilidade, a influência dessas informações incorporadas através do hiperparâmetro nas estimativas de credibilidade vai depender do quanto se acredita que essas informações sejam precisas.

Resultados solução II:

A Figura 4.4 apresenta as trajetórias dos valores simulados para μ , θ_1 e \tilde{Y}_1 , enquanto na Figura 4.5 se visualiza os gráficos de densidade de Kernel destes parâmetros. As interpretações são análogas a solução I.

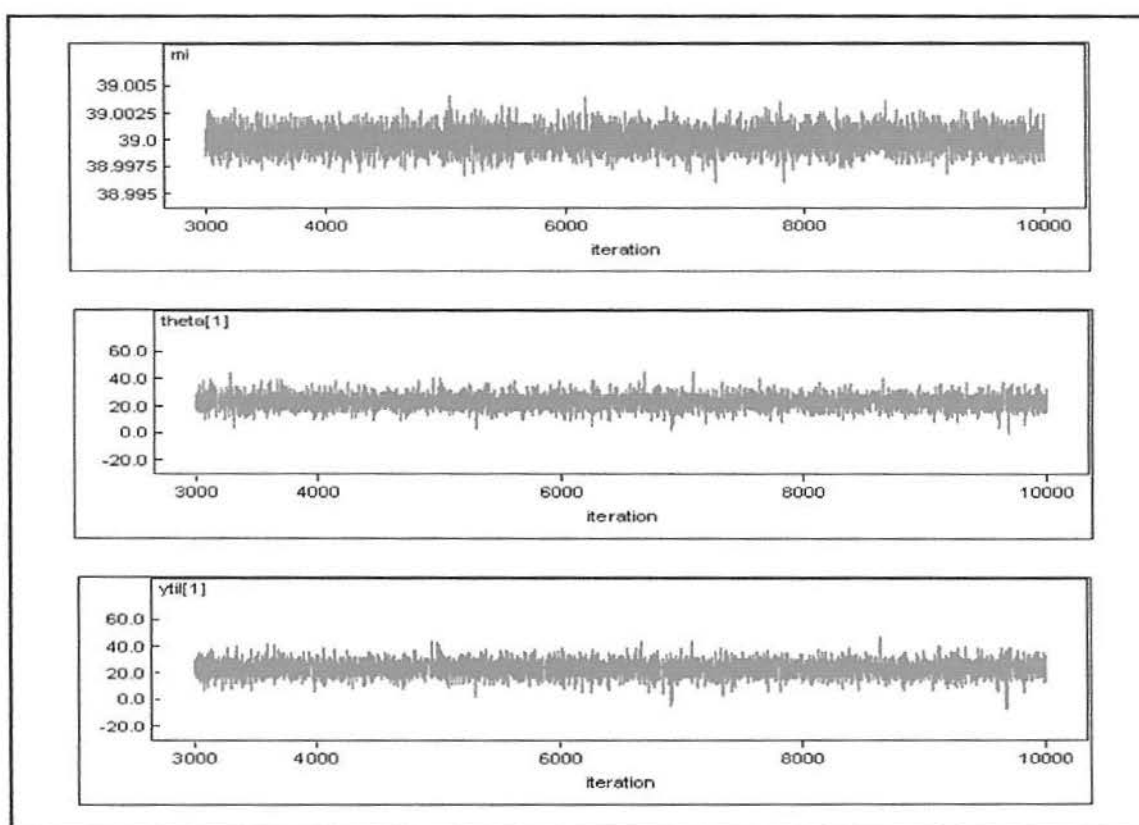


Figura 4.4 – Gráfico das trajetórias da simulação de μ , θ_1 e \tilde{Y}_1 , solução II

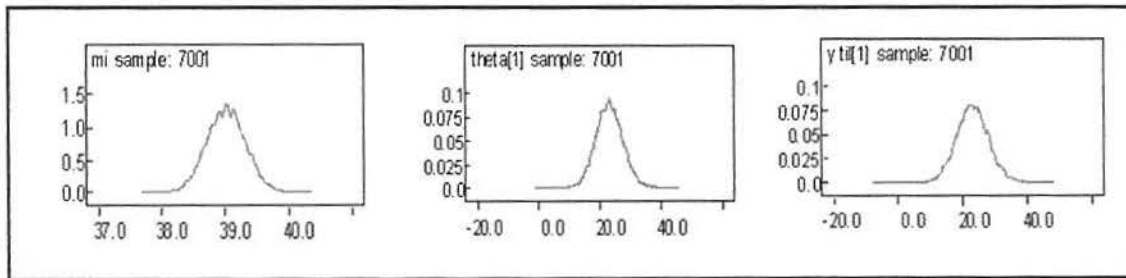


Figura 4.5 – Gráfico dos núcleos das densidades a posteriori de μ , θ_1 e \tilde{Y}_1 , solução II

A Tabela 4.6 apresenta os resultados descritivos a posteriori para os parâmetros do modelo. Novamente os resultados de \tilde{Y} serão utilizados para prever o número de sinistros de cada classe de risco para período cinco. Esta previsão, obtida a partir de simulação pelo Método Gibbs Sampling, é comparada com o valor real observado, com o valor esperado pela Tábua Biométrica AT-83, e com a Solução 1.

Parâmetros	media	D.P.	2.5%	97.5%	mediana
μ	39,10	0,310	38,400	39,700	39,000
σ_ε^2	0,0011	0,0008	0,0001	0,0010	0,0008
σ_θ^2	0,2336	0,0866	0,0974	0,4332	0,2230
θ_1	24,100	4,696	15,000	32,500	24,000
θ_2	58,800	4,151	53,300	68,800	59,100
θ_3	82,700	4,279	78,900	95,000	82,300
θ_4	103,100	5,453	95,700	117,400	102,900
θ_5	47,700	10,430	29,000	70,000	48,000
\tilde{Y}_1	24,100	5,192	13,000	33,300	24,000
\tilde{Y}_2	58,800	4,690	50,900	70,000	58,500
\tilde{Y}_3	82,700	4,851	75,900	97,000	83,000
\tilde{Y}_4	103,100	5,844	95,000	118,000	103,500
\tilde{Y}_5	47,700	10,610	29,000	70,700	47,900

Tabela 4.6 – Resultados a posteriori para os parâmetros do modelo, solução II

Na Tabela 4.7, se apresenta um quadro comparativo entre as duas soluções apontadas acima, através da aplicação da Teoria da Credibilidade sob enfoque Bayesiano, com a estimativa obtida pela Tábua Biométrica AT-83 e o valor real observado para o período cinco.

Pode-se constatar que, apesar dos resultados apontados pela solução II estarem muito próximas do valor observado para o número de sinistros do período cinco, a previsão para o número total de sinistros para este período agora é subestimada em relação ao valor observado. Enquanto a previsão aponta para um total de 316,4 sinistros no período, o número observado foi de 323.

Classe	Observado	AT-83	Solução I	Solução II
I	22,0	14,0	23,9	24,1
II	58,0	19,0	61,0	58,8
III	86,0	35,0	87,0	83,7
IV	106,0	84,0	107,0	103,1
V	51,0	41,0	52,0	47,7
Total	323,0	193,0	330,9	316,4

Tabela 4.7 – Comparativo entre valores preditos, observados e esperados pela Tábua AT-83 para o Ano 5

A subestimação apontada na previsão para o número total de sinistros obtida pela solução 2 pode ser entendida como o efeito da incorporação de informações a priori no modelo, através dos valores fixados para a hiperpriori de μ . Essa informação a priori incorporada ao modelo tinha como previsão um número total de 193 sinistros para o período 5. Assim sendo, é natural que ao conjugar essa informação com os dados observados na amostra, a previsão tenda a diminuir. Essa diminuição irá depender, como já mencionado, da precisão atribuída a distribuição de μ . Quanto maior for essa precisão, mais próximo de 193 deverá ser a previsão para o período 5. No entanto, essa discussão não será abordada neste trabalho, ficando como sugestão para trabalhos futuros.

De uma maneira geral, este exemplo ilustrativo de aplicação prática da Teoria da Credibilidade, sob um enfoque Bayesiano, procurou demonstrar a aplicabilidade da Técnica em dados reais de uma apólice. Observou-se sua eficácia na obtenção de previsões utilizando-se apenas dos dados observados, ou seja, com utilização de prioris não-informativas no modelo, assim como também a praticidade em incorporar informações subjetivas, ou alheias aos dados, a partir da atribuição de prioris informativas para os parâmetros desconhecidos do modelo. Esse é, sem dúvida, o principal ganho na utilização de métodos Bayesianos na Teoria da Credibilidade.

5. CONSIDERAÇÕES FINAIS

O objetivo deste trabalho foi, além de apresentar os principais fundamentos da Teoria da Credibilidade sob um enfoque Bayesiano, ilustrar a partir de um exemplo prático a aplicação da técnica em uma apólice de seguro de vida em grupo, através da implementação de métodos Bayesianos na Teoria da Credibilidade.

Os conceitos Bayesianos foram apresentados e aplicados aos modelos de Credibilidade de Bühlmann e Bühlmann-Straub. A partir dessa aplicação podem-se destacar alguns aspectos interessantes sobre a utilização dos métodos Bayesianos na Teoria da Credibilidade. Um destes aspectos se refere aos resultados obtidos da Inferência Bayesiana. As estimativas dos parâmetros desconhecidos do modelo são obtidas diretamente de uma distribuição de probabilidade, distribuição a posteriori, trazendo uma maior facilidade na interpretação dessas estimativas. Além disso, com uso de Gibbs Sampling e outras técnicas MCMC, o ajuste de modelos de Credibilidade mais complexos, quando, por exemplo, o número de parâmetros desconhecidos do modelo é muito grande, fica facilitado. Somando-se a isso, os métodos MCMC permitem a realização de inferências exatas para qualquer tamanho de amostra sem restrições dos cálculos assintóticos. E por último, ainda, a utilização de métodos Bayesianos permite a incorporação de informações a priori que se tenha sobre os parâmetros do modelo.

Mais especificamente sobre o exemplo de aplicação, deve-se reforçar a idéia de que o mesmo teve um intuito apenas ilustrativo, com o objetivo de demonstrar como a técnica, com implementação de métodos Bayesianos, pode ser utilizada em uma das principais tarefas da área de tarifação de seguros: a previsão. Assim, não houve preocupação em provar a eficácia desses métodos e nem compará-los com outros métodos existentes na literatura.

Dessa forma, no exemplo prático deste trabalho não foi considerado, por exemplo, o fato de que a suposição de normalidade para a variável utilizada na aplicação, número de

sinistros ocorridos em uma classe de risco, pode não ser a mais indicada. Isso porque as observações dessa variável são não-negativas, enquanto a distribuição normal pode assumir valores negativos. Além disso, pode haver também casos em que uma grande quantidade de classes não possua registros de sinistro em um determinado período, assumindo valor zero nestes períodos. Isso acaba definindo o ponto zero como um ponto que possui densidade diferente de zero. Assim sendo, seria natural se pensar em uma distribuição discreta, Poisson, por exemplo, para a variável número de sinistros de cada classe.

Como sugestão para estudos futuros, além de desenvolver extensões dos modelos para variáveis sem suposição de normalidade, seria interessante a abordagem da questão de como a precisão das informações a priori sobre os parâmetros deve ser atribuída e o quanto essa precisão atribuída as informações irá influenciar as estimativas obtidas.

Assim, este trabalho pode ser utilizado como uma fonte básica para consulta sobre a utilização de métodos Bayesianos aplicados a Teoria da Credibilidade, uma vez que foram apresentados os principais conceitos Bayesianos aplicados a esta Teoria, com utilização de um exemplo prático para ilustrar a implementação de métodos de estimação Bayesianos na Teoria da Credibilidade.

REFERÊNCIAS BIBLIOGRÁFICAS

BAYES, T. (1763). *An essay towards solving a problem in the doctrine of chances philosophical Transactions of the Royal Society*, 330-418. Reprinted, with biographical note by G.A Barnard, in *Biometrika* 45, 293-315 (1958).

FERREIRA, P. P. (2002) *Modelos de precificação e ruína para seguros de curto prazo*. Rio de Janeiro: Funenseg.

GILKS, W. R., RICHARDSON, S. & SPIEGELHALTER, D. J. (1996) *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.

LONGLEY-COOK, L.H. (1962) An introduction to credibility theory. *Proceedings of the casualty Actuarial Society*, New York, v. 49, p. 194-221.

MANO, C. M. C. A. B. (1998) *Melhoria da qualidade na tarifação de seguros: Uso de modelos de credibilidade*. Rio de Janeiro: Funenseg.

MEYER, P. L. (1983) *Probabilidade – Aplicações à Estatística*. Rio de Janeiro: LTC.

PEREIRA, F. C. (1998) *Teoria da Credibilidade: uma abordagem integrada*. Rio de Janeiro: Funenseg.

Anexo 1 - Definições básicas de Cadeias de Markov

Uma **cadeia de Markov** é um caso especial de um processo estocástico e é definida como uma sucessão de variáveis aleatórias $X_0, X_1, \dots, X_n, \dots$, tal que a distribuição de X_{n+1} dado X_n é independente de X_{n-1}, X_{n-2}, \dots , ou seja:

$$P[X_n = i / X_0, X_1, \dots, X_{n-1}] = P[X_n = i / X_{n-1}].$$

Estamos a supor que X_n designa o estado do sistema no período n e que em cada período n , X_n pode assumir N estados possíveis, $X_n \in \{1, 2, \dots, N\}$.

A cadeia de Markov diz-se **homogênea** se $P[X_n = i / X_{n-1}]$ não depende do período n , ou seja, suas probabilidades de transição de um estado para outro são invariantes.

Uma cadeia de Markov homogênea fica definida pelas probabilidades de transição condensadas na chamada matriz de transição (matriz quadrada):

$$P = [P_{ij}]_{i,j=1,2,\dots,N}$$

$$P_{ij} = P(X_{n+1} = j / X_n = i)$$

Além disso, uma cadeia de Markov diz-se **irredutível** se para qualquer par (i, j) existe uma probabilidade positiva de um processo que está no estado i atingir o estado j em algum período à frente, ou seja, $P_{ij}^{(n)} > 0$, para algum n . Isso equivale a dizer que cada estado da cadeia pode ser atingido a partir de qualquer outro estado em um número finito de iterações.

$$P_{ij}^{(n)} = P(X_{t+n} = j / X_t = i)$$

Para uma cadeia de Markov irredutível pode-se definir $\pi_j, j = 1, 2, \dots, N$, - proporção de períodos em que o processo se encontra no estado j no longo prazo, ou seja, quando $n \rightarrow \infty$ - como a **distribuição estacionária** da cadeia de Markov.

Pode-se demonstrar que $\pi_j, j = 1, 2, \dots, N$ existem e que são soluções únicas para o seguinte sistema de equações:

$$\pi_j = \sum_{i=1}^N \pi_i P_{ij}, j = 1, 2, \dots, N, \text{ onde } \sum_{j=1}^N \pi_j = 1.$$

No entanto, a existência de uma distribuição estacionária da Cadeia de Markov não é o suficiente para o propósito do Método MCMC. É necessário que as interações da cadeia converjam para a distribuição estacionária, pois se quer usar uma cadeia de Markov que, no limite, tenha a distribuição (distribuição estacionária) que se pretende simular.

Demonstra-se que se pode obter essa convergência se for considerada uma cadeia de Markov aperiódica.

Uma Cadeia de Markov irredutível diz-se **aperiódica** se não possuir estados absorventes, ou seja, se para algum $n \geq 0$ e para algum j , o seguinte ocorre:

$$P\{X_n = j / X_0 = j\} > 0 \text{ e}$$

$$P\{X_{n+1} = j / X_0 = j\} > 0$$

Se uma cadeia de Markov é irredutível e aperiódica, então se têm a seguinte relação entre a probabilidade de transição a n passos $P_{ij}^{(n)}$ e a distribuição estacionária π_j :

$$\lim_{n \rightarrow \infty} P\{X_n = j\} = \pi_j, j=1, 2, \dots, N.$$

Assim, concluí-se que, independentemente do seu estado inicial, na medida em que $n \rightarrow \infty$ a cadeia tende a depender cada vez menos deste estado inicial e que, além disso, a distribuição de probabilidade que fornece as probabilidades com que a cadeia se move de seu estado inicial i , $P_i^{(n)}$, e visita qualquer outro estado do espaço de estados, em n passos, irá convergir para a distribuição estacionária π_j .

Então, se o objetivo é gerar um vetor de variáveis aleatórias com distribuição de probabilidade π e dimensão k , pode-se utilizar uma Cadeia de Markov irreduzível e aperiódica com distribuição estacionária π . Assim sendo, inicia-se a Cadeia de Markov em qualquer ponto inicial e deixa-se a cadeia percorrer um número suficiente de períodos. As últimas K simulações constituem o vetor pretendido, pois à medida que n aumenta os pontos X_n gerados a partir da matriz de transição definida em (*) irão cada vez mais se parecer com amostras de π .

Assim sendo, se faz necessário definir a matriz de transição a ser utilizada para que a distribuição estacionária π seja a distribuição de interesse. Com esse objetivo muitas técnicas foram desenvolvidas a partir do algoritmo de Metropolis-Hastings.

Anexo 2 – Programação em R para exemplo de simulação

```
contador=100 # número de repetições
q=c(rep(0,contador))
matriz.diferença <- matrix(c(rep(0,(contador*5))), contador, 5)
matriz.diferença2 <- matrix(c(rep(0,(contador*5))), contador, 5)

for (c in 1:contador)
{
n= 2 #períodos observados em anos
k= 5 #número de classes da carteira

#parâmetros da simulação
sigma2= 25
tau= 10
mi= 100

media=c(0,0,0,0,0)
theta=c(0,0,0,0,0)
erro=c(0,0,0,0,0)

x=matrix(c(rep(0,(n*k))), n, k)
erro2=matrix(c(rep(0,(n*k))), n, k)
x2=matrix(c(rep(0,(n*k))), n, k)

for(j in 1:k)
{
erro[j]= mnorm(1,0,tau)
theta[j]= mi + erro[j]
}
  for (s in 1:n)
  {
    for (j in 1:k)
    {
      erro2[s,j]= mnorm(1,0,sigma2)
      x[s,j]= theta[j] + erro2[s,j]
    }
  }

J <- n # Número de anos
y <- x
```

```

ybarra <- c(rep(0,k))
for (i in 1:k) {ybarra[i] <- mean(y[,i])}

# Parametros das prioris
mi0 <- 0
sigma20 <- 100000
a1 <- 0.000001
b1 <- 0.000002
a2 <- 0.000001
b2 <- 0.000002

# Gibbs-sampling

m <- 10000 # Tamanho da cadeia
burn <- 3000
teta.sim <- matrix(rep(0, k*m), nrow=m, ncol=k)
sigma2e.sim <- rep(0, m)
sigma2teta.sim <- rep(0, m)
miteta.sim <- rep(0, m)
xtil <- matrix(rep(0, k*m), nrow=m, ncol=k)
sumsq <- matrix(rep(0, k*m), nrow=m, ncol=k)

#Valores Iniciais
teta.sim[1,1] <- 0
teta.sim[1,2] <- 0
teta.sim[1,3] <- 0
teta.sim[1,4] <- 0
teta.sim[1,5] <- 0
miteta.sim[1] <- 5
sigma2e.sim[1] <- 0.1
sigma2teta.sim[1] <- 0.1

for (i in 1:(m-1)) {
miteta.sim[i+1] <- rnorm(1, (sigma2teta.sim[i]*mi0 + sigma20*sum(teta.sim[i,]))/
(sigma2teta.sim[i]+k*sigma20), sqrt(sigma2teta.sim[i]*
sigma20/(sigma2teta.sim[i]+k*sigma20)))

sigma2teta.sim[i+1] <- 1/(rgamma(1, shape = (a1+0.5*k),
rate=(b1 + 0.5*sum((teta.sim[i,]-miteta.sim[i+1])^2))))
for(j in 1:k) {
teta.sim[i+1,j] <- rnorm(1,(J*sigma2teta.sim[i+1]*ybarra[j]/
(J*sigma2teta.sim[i+1]+sigma2e.sim[i])+sigma2e.sim[i]*miteta.sim[i+1]/
(J*sigma2teta.sim[i+1]+sigma2e.sim[i])), sqrt(sigma2teta.sim[i+1]*sigma2e.sim[i]/
(J*sigma2teta.sim[i+1]+sigma2e.sim[i])))}
}

```

```

for(j in 1:k) xtil[i+1,j] <- rnorm(1,teta.sim[i+1,j],sqrt(sigma2teta.sim[i+1]))
for (p in 1:k) {sumsq[i+1,p] <- sum((y[,p]-teta.sim[i+1,p])^2) }
sigma2e.sim[i+1] <- 1/(rgamma(1, shape = (a2+0.5*k*J),
rate=(b2 + 0.5*sum(sumsq[i+1,])))})
miteta <- miteta.sim[(burn+1):m]
mean(miteta)

ybarrab<- mean(ybarra)
vybarrab <- c(rep(ybarrab,5))
vteta.sim <- c(rep(0,k))
for (i in 1:k) {vteta.sim[i] <- mean(teta.sim[,i])}
diferença <- (ybarra - vteta.sim)^2
diferença2 <- (vybarrab - vteta.sim)^2
matriz.diferença[c,] <- diferença
matriz.diferença2[c,] <- diferença2
}
anos2a<-matriz.diferença
anos2b<-matriz.diferença2

anos3a<-matriz.diferença
anos3b<-matriz.diferença2

anos5a<-matriz.diferença
anos5b<-matriz.diferença2

anos8a<-matriz.diferença
anos8b<-matriz.diferença2

anos12a<-matriz.diferença
anos12b<-matriz.diferença2
boxplot(anos2a[,1],anos3a[,1],anos5a[,1],anos8a[,1],anos12a[,1],names=c("2 anos", "3
anos", "5 anos", "8 anos", "12 anos"))
title(main="Boxplot da diferença média entre Theta[1] e média(Y1)")

boxplot(anos2a[,2],anos3a[,2],anos5a[,2],anos8a[,2],anos12a[,2],names=c("2 anos", "3
anos", "5 anos", "8 anos", "12 anos"))
title(main="Boxplot da diferença média entre Theta[2] e média(Y2)")

boxplot(anos2a[,3],anos3a[,3],anos5a[,3],anos8a[,3],anos12a[,3],names=c("2 anos", "3
anos", "5 anos", "8 anos", "12 anos"))
title(main="Boxplot da diferença média entre Theta[3] e média(Y3)")

```

```
boxplot(anos2a[,4],anos3a[,4],anos5a[,4],anos8a[,4],anos12a[,4],names=c("2 anos", "3  
anos", "5 anos", "8 anos", "12 anos"))  
title(main="Boxplot da diferença média entre Theta[4] e média(Y4)")
```

```
boxplot(anos2a[,5],anos3a[,5],anos5a[,5],anos8a[,5],anos12a[,5],names=c("2 anos", "3  
anos", "5 anos", "8 anos", "12 anos"))  
title(main="Boxplot da diferença média entre Theta[5] e média(Y5)")
```

Anexo 3 – Programação em winbugs para exemplo prático

```
model;

const
N=5,
T=4;

Data in "sinistros.dat";

model
{
  mi ~ dnorm(39,10)
  sigmae ~ dgamma(0.01,0.01)
  sigmat ~ dgamma(0.01,0.01)

  for(i in 1 : N)
  {
    theta[i] ~ dnorm(mi,sigmae);

    for(j in 1:T)
    {
      sigc[j,i] <- sigmat*P[j,i];
      Y[j,i] ~ dnorm(theta[i],sigc[j,i])
    }

    ytil[i] ~ dnorm(theta[i],sigmat)
  }
}
```